6315c257a20d13521d898162ab9bd4479837ae031ecf33e5815991

Advances in Romance Linguistics

Publications in Language Sciences

Publications in Language Sciences in a series of monographs in theoretical linguistics and its more formal satellite fields of scientific inquiry. It is intended to provide an opportunity for scholars working within these areas to report rapidly on resaerch either terminated or in some definite stage of progress. Ideally, in these publications interest, inspiration, and openmindedness go hand in hand.

Ger J. de Haan
Leo Wetzels
Wim Zonneveld
editors

Other books in this series:

1. W. Zonneveld and F. Weerman *Linguistics in the Netherlands 1977-1979*
2. G. de Haan *Conditions on Rules*
3. T. Hoekstra , H. v.d. Hulst and M. Moortgat *Lexical Grammar*
4. P. Kiparsky *Explanation in Phonology*
5. A.J. Vitale *Swahili Syntax*
6. R.M. Martin *Logico-Linguistic Papers*
7. S. Dik *Functional Grammar*
8. L. White *Grammatical Theory and Language Acquisition*
9. S. Strauss *Lexicalist Phonology of English and German*
10. M.C. Shapiro and H.F. Schiffman *Language and Society in South Asia*
11. Simon C. Dik *Advances in Functional Grammar*
12. H. Bennis and W.U.S. van Lessen Kloeke *Linguistics in the Netherlands 1983*
13. M.P.R. Van den Broecke, V.J. van Heuven and W. Zonneveld *Sound Structures*
14. Harry M. Bracken *Mind and Language: Essays on Descartes and Chomsky*
15. Mieke Trommelen *The Syllable in Dutch*
16. Carlos Gussenhoven *On the Grammar and Semantics of Sentence Accents*
17. Hans Bennis and W.U.S. van Lessen Kloeke (eds.) *Linguistics in the Netherlands 1984*
18. Jaap van Marle *On the Paradigmatic Dimension of Morphological Creativity*
19. Larry M. Hyman *A Theory of Phonological Weight*
20. Frits Stuurman *Phrase Structure Theory in Generative Grammar*
21. H. Haider and M. Prinzhorn *Verb-Second Phenomena in Germanic Languages*
22. Martin Everaert *The Syntax of Reflexivization*
23. Leo Wetzels and Engin Sezer, (eds.) *Studies in Compensatory Lengthening*
24. Osvaldo Jaeggli and Carmen Silva-Corvalán (eds.) *Studies in Romance Linguistics*
25. Carol Neidle and Rafael A. Nuñez Cedeño (eds.) *Studies in Romance Languages*
26. John Anderson and Jacques Durand (eds.) *Explorations in Dependency Phonology*
27. Frans van Coetsem *Loan Phonology and the Two Transfer Types in Language Contact*

David Birdsong and
Jean-Pierre Montreuil (eds.)

Advances in Romance Linguistics

1988
FORIS PUBLICATIONS
Dordrecht - Holland/Providence - U.S.A.

Published by:
Foris Publications Holland
P.O. Box 509
3300 AM Dordrecht, The Netherlands

Sole distributor for the U.S.A. and Canada:
Foris Publications USA, Inc.
P.O. Box 5904
Providence RI 02903
U.S.A.

Sole distributor for Japan:
Toppan Company, Ltd.
Shufunotomo Bldg.
1-6, Kanda Surugadai
Chiyoda-ku
Tokyo 101, Japan

CIP-DATA

Advances

Advances in Romance Linguistics / David Birdsong and Jean-Pierre Montreuil (eds.). -
Dordrecht [etc.] : Foris. - (Publications in Language Sciences ; 28)
ISBN 90 6765 344 6 bound
ISBN 90 6765 345 4 paper
SISO roma 837 UDC 804
Subject heading: Romance linguistics.

ISBN 90 6765 344 6 (Bound)
ISBN 90 6765 345 4 (Paper)

© 1987 Foris Publications - Dordrecht

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Printed in The Netherlands by ICG Printing, Dordrecht.

Contents

Introduction vii

Marianne Adams
Parametric Change: Empty Subjects in Old French 1

James Blevins
Restructured Compounds and Romance Causatives 17

Denis Bouchard
En-Chain 33

Becky Brown
Problems with Past Participles Agreement in French and Italian Dialects 51

Thomas D. Cravens
Consonant Strength in the Romance Dialects of the Pyrenees 67

Jorge M. Guitart
The Case for a Syntax-dependent Post-lexical Module in Spanish Phonology 89

Paul Hirschbühler
The Middle and the Pseudo-Middle in French 97

Osvaldo A. Jaeggli
ECP Effects at LF in Spanish 113

Richard D. Janda and Terrell A. Morgan
El acentó dislocadó – pues cantadó – castellanó: On Explaining Stress-Shifts in Song-Texts from Spanish (and Certain Other Romance Languages) 151

John Earl Joseph
First-Person Singular Inversion in French 171

Marie-Odile Junker
Transitive, Intransitive and Reflexive Uses of Adjectival Verbs in French 189

Christine Kamprath
Morphological Constraints on Epenthesis and Adjunction: Paradigmatic or Cyclic? 201

Mark Vincent LaPolla
Clitic Movement in Spanish and the Projection Principle 217

Claire Lefebvre
Past Participle Agreement in French: Agreement = Case 233

Juliette Levin
Constraints on Rime-internal Syllabification in French: Eliminating Truncation Rules 253

France Martineau
Variable Deletion of *que* in the spoken French of Ottawa-Hull 275

Donna Jo Napoli
On Predication and Identity within NP's 289

Rafael A. Nuñez Cedeño
Structure-preserving Properties of an Epenthetic Rule in Spanish 319

Marianne Phinney
Resetting the Parameters: Acquiring Spanish as L2 337

Yves Roberge
Clitic-Chains and the Definiteness Requirement in Doubling Constructions 353

Donca Steriade
Gemination and the Proto-Romance Syllable Shift 371

Margarita Suñer
Quantifiers and Direct Object-Doubling 411

Marie-Thérèse Vinet
Implicit Arguments and Control in Middles and Passives 427

Introduction

The present volume consists of selected papers from the Sixteenth Linguistic Symposium on Romance Languages, held March 6-8, 1986, at the University of Texas at Austin. This forum, commonly referred to as LSRL, was initiated at the University of Florida in 1971 and has since been convened annually at various university sites in Canada and the United States. The University of Texas is pleased to have hosted both the fourth and the sixteenth meetings of the Symposium.

Romance Linguistics is a burgeoning field. Ample evidence of this trend is seen in professional publications, doctoral dissertations, and university course offerings. The field's strength has been further attested in the quantity and quality of scholarly activity in professional language societies. Among these, the LSRL has assumed a position of undisputed prominence. Over the last several years, participants in LSRL have represented not only U.S. universities, but institutions across Europe, South America, and Canada as well. Not surprisingly, competition for the presentation of papers at the Symposium has been vigorous. LSRL XVI was no exception: of the papers reviewed for presentation, fewer than one-fourth appear in this volume.

The title, ADVANCES IN ROMANCE LINGUISTICS, properly implies a collection of state-of-art research in the field. However, it is not meant to suggest a thorough, representative sampling of current methods, frameworks, and topics. Instead, emphasizing the notion of **advances**, the Editors and outside reviewers value the selections in this volume for their enhancement of the understanding of Romance languages. As a by-product of this emphasis (and not as the result of any professional bias), a disproportionate number of papers herein address various aspects of syntax. Considerably more evenness is visible, however, in the distribution of the Romance languages under investigation. While French and Spanish understandably predominate, Italian, Portuguese, Catalan, and several other variates of Romance are also represented. For whatever its balances or imbalances, ADVANCES IN ROMANCE LINGUISTICS is intended to capture the spirit and sophistication of linguistic inquiry in the mid-1980's, and to project an agenda for Romance research in the years to come.

We would like to express our gratitude to Dean Robert King of the College of Liberal Arts at the University of Texas, whose enthusiasm as a linguist and whose generosity as a dean made possible the hosting of LSRL XVI in Austin. For logistical and moral support, we sincerely thank Jean-Pierre Cauvin, Chair of the Department of French and Italian, as well as the departmental staff. Our thanks extend to the many anomymous reviewers of conference papers and to several colleagues at UT, particularly Marta Luján and Denis Bouchard, whose expertise was called upon before, during, and after the Symposium. We also appreciate generous donations of time and energy by graduate students in French Linguistics, Romance Linguistics, Linguistics, and Foreign Language Education. Finally, we would like to acknowledge the invaluable assistance of Leo Wetzels and Foris Publications in bringing this book to print.

James W. Harris is a true friend of the Linguistic Symposium on Romance Languages. Over the years he has been a member of the LSRL Executive Committee, reviewer of conference paper abstracts and, on two occasions, keynote speaker. The successes and progress of LSRL are due in no small part to Jim's contributions.

LSRL is not unique in its indebtedness to Jim Harris: for nearly two decades, the community of Romance researchers as a whole has profited from his scholarship. It is no exaggeration to say that Jim's work constitutes a standard by which to measure quality and creativity in Romance Linguistics.

It is fitting, then, that a book entitled ADVANCES IN ROMANCE LINGUISTICS should be dedicated to James Harris. Thanks for everything, Jim.

Marianne Adams

Parametric Change: Empty Subjects in Old French

1.0 Introduction

This paper combines traditional Romance philology with formal syntax in an effort to explain the change from prodrop to nonprodrop in Old French. As a consequence, a new theory of prodrop emerges.

Old French (OF) of the 12th and 13th centuries presents a form of prodrop and inversion unlike that of the modern prodrop languages studied. The facts of OF, when analyzed within the theory of generative grammer, lead us to new insights into the relationship between prodrop and inversion. I will argue that the phenomenon of prodrop is not itself a parameter of universal grammer (UG) but reduces rather to directional government and rich agreement. This approach helps explain the distribution of **pro** across languages and within a given language, and it helps explain the change from a grammer which licenses **pro** to one which does not.

2.0 Old French

First let us look at the relevant facts of OF, given in (1), (2), and (3). These facts are well documented and have been discussed by many scholars of OF.[1]

2.1. Prodrop

Observe first of all that null subjects typically occur only in main clauses in OF, seldom in complement causes. This is exemplified in (1):

(1) a. Si firent _____ grant joie la nuit.
So made (they) great joy that night.
(Robert de Clari XII)

b. Einsi corurent ____$_i$ par mer tant que il$_i$
Thus ran (they) by sea until they
vindrent à Cademelée.
came to Cadmée. (Villehardouin XXV)

c. Et se feroient ______$_i$ si durement
And each other struck (they) so hard
des espées que çou estoit grans mervelle
with swords that it was a great marvel
que il$_i$ ne s'entr'ocioient.
that they not each other kill.
(**Roman du Graal**, p. 231)

Pragmatic factors cannot explain this distribution. Notice in (1b) and (1c) that the subject is the same in the upper and lower clause. Purely pragmatic or functional considerations would lead us to expect the reverse order of empty and lexical subjects than that which actually occurs. Verbal inflection cannot account for the main/subordinate clause asymmetry either. Inflection was rich in OF and usually distinguished all six persons. Crucially, the use of inflections showed no asymmetry of the sort we see in (1b) and (1c).

2.2 Germanic Inversion

Second, OF shows the Germanic type of inversion but not the Romance type. Germanic inversion, or Subject-Aux Inversion (SAI), is limited to main clauses and preposes only the inflected verb or auxiliary (V$_I$); it yields the order VSO. Thus OF has the kind of inversion English has but without do-support. Examples are given in (2):

(2) a. Or li doint Dex santé.
Now to-him give God-NOM health.
"Now God give him health."
(**La Mort le Roi Artu** 41)

b. Einsint aama la damoisele Lancelot.
Thus loved the young lady Lancelot-ACC.
"Thus the young lady loved Lancelot."
(**La Mort le Roi Artu** 38)

c. Ha, amis, dont est ce venus?
Ah, friend, whence is this come?
(**La Chastelaine de Vergi** 755)

d. Qui as tu donc aamé?
Whom have you then loved? (**Eneas** 8497)

Romance free inversion, on the other hand, the kind we find in Italian and Spanish, applies in both main and subordinate clauses and shows no asymmetries of the sort we have seen for OF. Under this type of inversion the subject appears at the end of VP and the verb does not move; the order it yields is VOS. Thus in Italian, for example, (2b) and (c) would translate as (e) and (f), with the subject rather than the VP to the right:

e. Cosi amò Lancelot la signorina.
f. Da dove è venuto questo?

Romance inversion is atypical of OF and is rarely attested.

Note that the inversion we see in (2) is not the kind discussed in Kayne and Pollock (1978) and known as French Stylistic Inversion. Stylistic Inversion is like Romance inversion in that it moves the subject to the right of VP.

2.3 V2 Effects

Third, OF obeyed a verb-second (V2) constraint. As in German, the V2 effects of OF show up primarily in main clauses. In main clauses almost any constituent could come first, but V_I always came second. The examples in (1) and (2) show OF V2 effects.

Subject pronouns had not yet cliticized in OF. As independent arguments separate from the verb, they could satisfy the V2 requirement. This is shown in (3):

(3) a. Il l'a bien moustré a ceste assemblée.
He it-has well shown at that tournament.
(**La Mort le Roi Artu** 34)

b. Ge meïsme sui si confus.
I myself am so confused.
(cited by Einhorn 1974, p. 69)

(Cf. ModF: Moi-même, je suis ...)

Object pronouns, unlike subject pronouns, had already become clitics by the time of the earliest OF documents. They thus could not 'count' for V2 and did not therefore begin sentences (for example (2a) and (3a)).

The cliticization of subject pronouns in OF involved both the reanalysis of prodrop to nonprodrop and changes in the rhythm and intonation of phrases and sentences. The rhythm of OF was unlike that of Modern French (ModF). The phrase was not oxytonic as it is in ModF, and individual words had greater independence and could bear their own accent as they can in English and German. When subject pronouns began to cliticize to the verb, they continued to appear initially. By this time, however, violations of the V2 constraint were becoming common and null subjects infrequent. These changes, I believe, were closely related, and work in progress is devoted to sorting out some of the cause and effect relationships involved. It is significant that neither Spanish nor standard Italian has subject clitics, but that the Northern Italian dialects do. These dialects patterned in many ways like OF in their medieval phase, showing SAI, V2 effects, and an asymmetry between independent and dependent clauses with respect to empty subjects.[2] We return to these matters briefly below.

2.4 Head first

Finally, OF, like Romance in general but unlike German and Dutch, was a head first language with a basic SVO word order. This order appeared regularly in subordinate clauses, as we saw in (1), and often in main clauses when the subject, nominal or pronominal, served to satisfy the V2 constraint (see the examples in (3)).

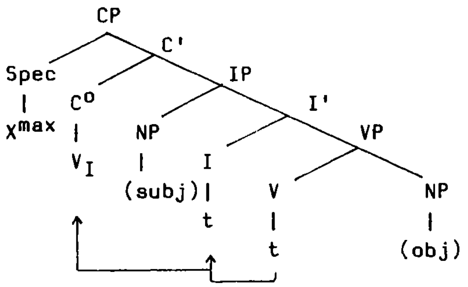
3.0 Analysis of V2 in OF

What stands out from the above set of facts is the asymmetry between independent and dependent clauses with respect to prodrop and V2 effects. We expect this for V2, which typically shows up only in main clauses, but the asymmetry with respect to prodrop is unexpected and unaccountable under current theories.

Before I suggest a different approach to prodrop, let's look at a formal derivation of V2 effects in OF. This is given in (4). The representation in (4) is obtained by adopting Thiersch's (1978) approach to V2 in German together with the more articulated structure of Comp suggested in Chomsky (1985). Note that unlike German and Dutch, both SOV languages the derivation for OF must begin with an underlying SVO order:

(4) V2 Effects in OF

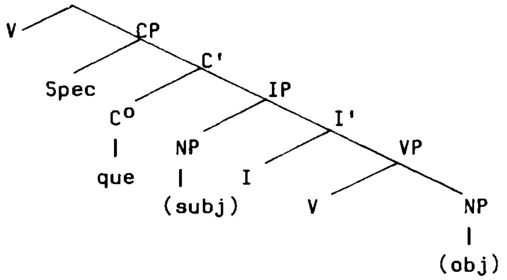
X^{max} = NP, PP, AdvP, etc.



In (4), the verb first raises to Infl where it picks up its inflection; then the combined form, V_I, moves to Comp. (Alternatively, Infl first moves to Comp, forcing the verb to join it (cf. Sproat 1985).) To maintain V2 effects, an independent consituent must precede the preposed V_I. In OF this can be almost any maximal projection: the subject, an object, a prepositional phrase, adjective phrase, adverbial phrase, WH phrase, quantifier phrase, or an entire clause. OF provides ample examples of all of these possibilities. The initial constituent either represents the theme of the sentence, or if nothing is thematicized, an adverbial of some sort, often one virtually devoid of meaning, serves to satisfy the V2 requirement (Herman 1954).

The absence of V2 effects in complement clauses follows from this analysis. In complement clauses the complementizer fills the verb's landing site, preventing it from preposing and forcing it to remain S-internal. This is shown in (5):

(5) Complement clauses in OF



4.0 Directional Government and the Theory of Prodrop

If the theory of prodrop is to take into account the facts of OF, it will need to be revised. The theory developed here can account for the cluster of properties presented above and for the transition from the OF to the ModF system, which is neither prodrop nor verb-second.

We saw that although OF licenses null subjects, it is normally only in main clauses that they are attested. Likewise, it is only in main clauses that V_I-fronting and V2 effects occur. In complement clauses, where V_I-fronting typically does not occur, a lexical pronoun regularly does. The illustration in (4) gives the context in which empty subjects occur; (5), that in which lexical pronouns show up.[3]

A directional government requirement on **pro** can account for these facts. This means that in order for Infl to license **pro**, it must be able to govern it in the proper direction. Direction of government, or canonical government, is widely assumed to be a parameter, the value of which is identified by the order in which a verb governs its complements: government to the right in a head first language, to the left in a head final language.[4] One way of obtaining canonical government of the subject by Infl in head first languages is by moving the subject to the end of VP. Another way is by moving the inflected verb to the left of the subject. Since the former, Romance inversion, was unavailable in OF, such government could not be obtained as it is in Italian and Spanish by postposing the subject. Canonical government was obtained in OF, however, when V_I preposed. This it did in main clauses, but it could not do so in complement clauses for the reasons given. The characteristic asymmetry of OF - empty subjects in main clauses, lexical subjects in complement clauses - thus follows straightforwardly from the requirement that **pro** be governed canonically.[5]

The theory of prodrop proposed in (6) incorporates canonical government:

(6) **Theory of Prodrop**
 a. The position and content of **pro** must be identified.
  i. **Pro's** position is identified by a governing head.

ii. **Pro's** content is identified by coindexation with the proper features.

b. Government of **pro** (6ai) must be in the canonical direction.

(Later we extend directionally from the theory of **pro** to the theory of government.) In the case of subject **pro**, Infl is the governing head which identifies **pro's** position, and features within Infl (agreement features (Agr)) identify **pro's** content. Hence Infl serves both as head governor and as antecedent of subject **pro**. When Infl identifies position, (6ai), but a minimum of content, (6aii), only expletive **pro** is available; that is, **pro** is identified, but receives no features.

Our (6a) is similar to Rizzi's (1985) theory of prodrop.[6] But Rizzi takes no account of (6b). For Rizzi, the prodrop parameter is whether or not a governing head can license **pro**, suggesting that languages are prodrop or non-prodrop at random according to whether or not they stipulate that such licensing is permissible. Let me suggest the possibility, however, that prodrop is not a parameter per se, but a descriptive generalization reducible to other properties, namely, directional government and feature identification, (6a) and (6b). If this is so, then **pro** should be available to any grammar which satisfies these conditions. We return to this shortly.

4.1 Italian and Spanish

The theory outlined in (6) predicts the position of **pro** in languages like Italian and Spanish, something left open under standard theories. This can be seen in (7), where if canonical government is to hold, **pro** must be postverbal:

(7) a. Vuole caffé col latte **pro**.
b. Quiere café con leche **pro**.
wants-3s coffee with milk

In certain contexts Spanish and Italian allow V_1-preposing. When they do, canonical government holds of the pre-VP subject position and **pro** may occur there. In Spanish the verb preposes in the context of WH-movement (Torrego 1984). This is illustrated in (8):

(8) No sabía qué querían.
not knew (I) what want (they)
"I didn't know what they wanted."
$\ldots[_{CP}\ \text{qué}_j\ [_{C'}\ \text{querían}_i\ [_{IP}\ \text{pro}\ t_i\ t_j]]]$

In Italian preposing occurs in the Aux-to-Comp structures discussed in Rizzi (1982) and exemplified in (9):

(9) Essendo molto improbabile che Carlo ci aiuti ..
being (it) very improbable that Carlo us help ..
"It being very improbable that Carlo help us .."
$[_{C'}\ \text{Essendo}_i\ [_{IP}\ \text{pro}\ t_i\ \text{molto improbabile..}]]$

In Italian only gerunds and, at a marked stylistic level, infinitives prepose to Comp. Since these forms lack agreement features, only expletive **pro** can be identified in this context in Italian.

The theory of **pro** advocated here lends support to the arguments of Borer (1985) and Travis (1984) that when the subject is post-VP there is no need for a pre-VP empty expletive and that VOS order may be base generated directly. This follows straightforwardly if directional government is to be maintained as a general condition on **pro**, referential and expletive alike. Hence (10a) but not (10b) is a possible configuration:

(10) a. $[_{IP}\ \text{VP}\ \text{pro}_{+\theta}\ ]$
b. $[_{IP}\ \text{pro}_{-\theta}\ \text{VP}\ \text{pro}_{+\theta}]$

4.2 Why are German and Dutch not prodrop languages?

An interesting consequence of our theory is that it helps answer a question often asked; namely, given their reasonably rich set of inflections (four distinctions), why are German and Dutch not prodrop languages? The answer follows from the fact that these languages are head last, but they have Comp on the left. So movement to Comp can never result in a canonical government configuration in German and Dutch, however rich their verbal inflections.

4.3 Directional Government

In (6b) I assumed only that direction is a condition on **pro**. The theory of prodrop can be generalized by taking the obvious next step and assuming that directionality is a condition on government itself, part of the theory of government (cf. Horvath 1981; Stowell 1983, 1985). This more general approach is to be preferred on theoretical grounds. Its adoption means that (6b) need no longer be specified as part of the theory of prodrop. As part of the theory of government itself it will constrain any government relation. (11) informally defines directional government:[7]

(11) **Directional Head Government**
The direction in which a head (an X^0 category) governs is a parametric choice. A head can only govern in the chosen direction.

We do not need to specify head government in (11) if the antecedent relation in (6aii) is not a government relation but rather a type of binding or identification condition which does not require government.[8] If we accept this, then (11) can be generalized to (12):

(12) **Directional Government**
Direction of government is a parametric choice. Government only takes place in the chosen direction.

Now the theory of prodrop can be simplified from that given in (6) to the revised version in (13):

(13) **Theory of Prodrop** (revision of (6))
The position and content of **pro** must be identified.
a. **Pro**'s position is identified by a governing head.
b. **Pro**'s content is identified by coindexation with the proper features.

4.4 Summary

I have tried to show that the phenomenon of prodrop need not be construed as the consequence of a special parameter of universal grammar. Rather, the availability of empty subjects may follow more simply from a conjunction of independent properties, to wit, directional government of **pro**'s position by Infl and identification of **pro**'s features by Agr. If this is correct, then null subjects should be available to any language which can satisfy these conditions. Both conditions are necessary to license **pro**, and syntactic change may occur when either one is altered.

Next we will see that the absence of Romance inversion in a head first language helps account for the change from prodrop to nonprodrop despite the presence of rich verbal inflection.

5.0 Loss of Prodrop in OF

We cannot here go into all of the interacting factors which led to the loss of prodrop in OF. I remarked earlier that feature identification was satisfied and cannot account for the distribution of lexical and empty subjects that we find. The traditional view, under which the phonological erosion of endings leads to a state of ambiguity and confusion which in turn forces the use of subject pronouns, cannot therefore be correct and probably puts the cart before the horse: parts of words are not lost when grammatical principles require them; they are lost rather when they are superfluous.

My view is that OF lost the ability to license **pro** because it lost V2 effects, and it lost V2 effects because its underlying word order was SVO. This order, it will be recalled, showed up consistently in complement clauses and frequently also in main clauses when the subject was thematicized. (Because the subject often constitutes the theme of a sentence, German and Dutch main clauses too often end up SVO.) But learners have no direct access to adult grammar and they draw conclusions strictly on the basis of what they hear (see H. Andersen 1973; Lightfoot 1979; Lightfoot and Hornstein 1981; Hyams 1983). Therefore, the assumption that learners misconstrued the derived SVO (V2) order of main clauses in OF as basic and that this mistake was reinforced by the consistent SVO order of complement clauses seems very plausible. (14) illustrates what I have in mind:

(14) $[S\ [V_I\ [_{IP}\ t\ t\ O\ ]]] \longrightarrow [_{IP}\ SVO\ ]$

Reanalysis of this sort would be unlikely to happen in German and Dutch, for in these languages the main clause order is never the same as that of the embedded clause.

The strict V2 effects of early OF began to break down in the late OF and Middle French periods. Initial prepositional phrases, direct and indirect object NPs, and so on, which earlier invariably induced V_I-fronting (XVS) increasingly failed to, giving rise to frequent XSV orders. This breakdown had obvious consequences for prodrop. Under our theory V_I-preposing was the only means OF had of obtaining directional government; without it, **pro** could no longer be licensed. Prodrop and V2 effects coexisted with the new system for a time as formal, increasingly frozen, literary alternatives. But eventually the written grammer caught up with the spoken and by the end of the 15th century, **pro** and V2 were largely a thing of the past.

5.1 Medieval Northern Italian Dialects

A survey of medieval Romance texts by Vanelli, Renzi, and Benincà (to appear) lends support to our theory of prodrop and change. This study shows that a chain of events similar to that which took place in OF took place also in the Northern Italian, Franco-Provençal, and Swiss Romande dialects. All of these dialects show in their medieval phase V2 effects and prodrop in main clauses and consistent SVO order in complement clauses. These dialects lost V2 effects by the end of the 15th century. Significantly, it is these dialects which today have obligatory subject clitics. As I have argued elsewhere (Adams 1985), these clitics indicate that at some point in their history the dialects were nonprodrop.

The approach to change that I am advocating supports the argument that head government must hold at S-structure (Stowell 1985) or surface structure (Jaeggli 1985; Weinberg, et al. 1986). Head government at LF, as when Infl raises to Comp at that level (Stowell 1981), comes too late. It is too late because canonical government must hold in the primary data if learners are to draw conclusions about their grammar based upon it. If canonical government holds at S-structure, it will normally hold at surface structure since even if superficial stylistic rules occasionally alter word order, such variations are by their nature exceptional and do not undermine generalizations.

5.2 Loss of Prodrop in Brazilian Portuguese

OF provides empirical evidence for Safir's (1982) claim that prodrop and Romance free inversion are separate properties. It is evident from our analysis, however, that the properties are closely related: Romance inversion is one way of licensing **pro**; it is thus no accident that they often cooccur.

The loss of both Romance inversion and V_I-preposing should, then, have a direct effect on the loss of **pro**, if our theory is correct. Brazilian Portuguese provides interesting support in this regard.[9] It has lost free inversion and unlike Continental Portuguese it lacks V_I-preposing. Since it thus has no way to canonically govern a subject, it should be nonprodrop. In effect, Kempchinsky (1985) argues that Brazilian Portuguese is losing the prodrop parameter. Her argument is based on the observation that Brazilian Portuguese has lost a number of properties commonly associated with prodrop. First, as noted, it no longer licenses free inversion. Second, it regularly uses subject pronouns without emphatic force (something far less common in Italian and Spanish), and in the third person singular and plural it actually prefers them. Third, as illustrated in (15), Brazilian Portuguese shows no preference for a disjoint reading, whereas Italian and Spanish prefer one:

(15) Ele$_i$ diz que êle$_{i,j}$ està doente.
He says that he is sick.

Fourth, it shows subject-object asymmetries typical of nonprodrop languages (Zubizarreta 1982 and Kempchinsky 1985 argue that Portuguese has a **que/que** rule equivalent to the **que/qui** rule of French). Finally, it has a reduced set of verbal inflections from that of Continental Portuguese.[10]

Notice that these changes in Brazilian Portuguese are taking place despite its rich heritage of prodrop literature and despite the fact that it is geographically surrounded by speakers of an unambiguously prodrop language.

6.0 Conclusion

The pattern of prodrop found in OF suggested a revision in the theory of **pro**. The theory proposed removes the phenomenon of prodrop from the set of universal parameters and shows that it very probably can be reduced to directional government and feature identification. We saw this approach to have a number of interesting consequences well beyond its ability to account for the facts of OF. Synchronically, it helps explain the distribution of **pro** within languages and across languages. Diachronically, our theory helps explain the change from a system which licenses **pro** to one which does not.

NOTES

* I wish to thank Tim Stowell, Nina Hyams, Richard Kayne, and David Lightfoot.

[1] See Foulet (1928), Franzen (1939), Herman (1954), Price (1966), Thurneysen (1892), Wagner (1974), Wartburg (1946), Zwanenburg (1979), and Benincà (1984). The data presented in (1)-(3) are remarkably consistent in OF prose; in verse, to be sure, there is more variation, but the generalizations hold well nonetheless.

[2] See Vanelli, Renzi, and Benincà (to appear).

[3] The insight that null subjects occur in the context of inversion, hence in main but not in subordinate clauses, goes back to Foulet (1928), Franzen (1939), Herman (1954) and others. The observation that OF was a V2 language goes back to Thurneysen (1892) and has continued to be widely accepted. Vanelli, Renzi, and Benincà (to appear) and Benincà (1984) observe that certain other medieval Romance languages behaved like OF. Following Thiersch (1978), they insightfully reinterpret these observations in a manner similar to (4) and (5).

[4] See Horvath (1981), Kayne (1984), Stowell (1983, 1985), and Adams (1985) for other arguments in favor of directional government.

[5] V_I-fronting also occurs in the context of **yes/no** questions in OF. Here no constituent need precede V_I, yet **pro** is nonetheless available because the preposed verb canonically governs it:

(i) Savès ____ pour coi j'ai men abit cangiét?
Know (you) why I have my attire changed?
(Cited by Foulet 1928, 314)

[6] Rizzi (1985) incorporates cases of object **pro** into the theory of prodrop by extending the set of possible governing heads from Infl to V and P and the set of identifying features from Agr features to θ-roles. The cases he discusses appear to be compatible with both (6a) and (b). Rizzi's theory can be summarized as follows:

(i) **Pro** must be licensed by a governing head of type X. This is parametric.
(ii) Pro must be identified by features of X. This recovers the content of **pro**.

[7] The generalization requires that the head government

relation be distinguished from the antecedent relation, but this distinction, it seems, needs to be made anyway. Kayne (1981) was perhaps the first to articulate it by requiring that an empty category be both lexically governed (by X^{0}) **and** within a percolation projection of its antecedent. Jaeggli (1982) specifically split proper government into c-government and s-government, essentially antecedent government and head government. This approach has been developed more recently in Jaeggli (1985), Rizzi (1985), Stowell (1985), Travis (1984), Weinberg, et al. (1986), and others. Items (6) and (13) incorporate it.

[8] See Stowell (1985) and Weinberg, et al. (1986) for arguments to this effect.

[9] The Portuguese data are drawn from Chao (1980), Zubizarreta (1982), and Kempchinsky (1985). I am grateful to Paula Kempchinsky and Carlos Quicoli for helpful discussion.

[10] Here are the Continental and standard Brazilian Portuguese paradigms, (i) and (ii) respectively:

(i)	eu lavo	nós lavamos
	tu lavas	vós lavais
	êle/ela lava	êles/elas lavam
(ii)	eu lavo	nós lavamos
	tu/êle/ela lava	vós/êles/elas lavam

REFERENCES

Adams, M. 1985. Parametric Change. Ms. UCLA.

Andersen, H. 1973. Abductive and Deductive Change. Language 49.765-93.

Benincà, P. 1984. Un'Ipotese sulla sintassi delle lingue romanze medievali. Padua: Quaderni Patavini di Linguistica.

Borer, H. 1983. I-subjects. Ms. UC Irvine.

Chao, W. 1980. PRO-drop languages and nonobligatory control. University of Massachusetts Occasional Papers in Linguistics 7.46-74.

Chomsky, N. 1985. Barriers. Ms. MIT.

Einhorn, E. 1974. Old French: A Concise Handbook. Cambridge University Press.

Foulet, L. 1928. Petite syntaxe de l'ancien français. Paris: Champion.

Franzen, T. 1939. La syntaxe des pronoms personnels sujets en ancien français. Uppsala: Almqvist.

Herman, J. 1954. Recherches sur l'ordre des mots dan les plus anciens textes français en prose. Acta Linguistica Hungarica 4.69-94, 351-382.

Horvath, J. 1981. Aspects of Hungarian Syntax and the Theory of Grammar. Ph.D. Dissertation, UCLA.

Hyams, N. 1983. The Acquisition of Parameterized Grammars. Ph.D. Dissertation, CUNY.

Jaeggli, O. 1982. Topics in Romance Syntax. Dordrecht: Foris.

Jaeggli, O. 1985. On Certain ECP Effects in Spanish. Ms. USC.

Kayne, R. 1981. ECP Extensions. Linguistic Inquiry 12.93-133.

Kayne, R. 1984. Connectedness and Binary Branching. Dordrecht: Foris.

Kayne, R. and J.-Y. Pollock. 1978. Stylistic Inversion, Successive Cyclicity, and Move NP in French. Linguistic Inquiry 9.595-621.

Kempchinsky, P. 1985. Brazilian Portuguese and the Null Subject Parameter. Mester 13.2. Department of Spanish and Portuguese, UCLA.

Lightfoot, D. 1979. Principles of Diachronic Syntax. Cambridge University Press.

Lightfoot, D. and N. Hornstein. 1981. Explanation in Linguistics. Longman.

Price, G. 1966. Contribution à l'étude de la syntaxe des pronoms personnels sujets en ancien français. Romania 87.476-504.

Rizzi, L. 1982. Issues in Italian Syntax. Dordrecht: Foris.

Rizzi, L. 1985. Null Objects in Italian and the Theory of pro. Ms. MIT.

Safir, K. 1982. Syntactic Chains and the Definiteness Effect. Ph.D. Dissertation, MIT.

Sproat, R. 1985. Welsh Syntax and VSO Structure. Natural Language and Linguistic Theory 3.172-216.

Stowell, T. 1981. Origins of Phrase Structure. Ph.D. Dissertation, MIT.

Stowell, T. 1983. Subjects Across Categories. The Linguistic Review 2.3. Dordrecht: Foris.

Stowell, T. 1985. Null Operators and the Theory of Proper Government. Ms. UCLA.

Thiersch, C. 1978. Topics in German Syntax. Ph.D. Dissertation, MIT.

Thurneysen, R. 1892. Die Stellung des Verbums im Altfranzösischen. Zeitschrift fur romanische Philologie 16.

Torrego, E. 1984. On Inversion in Spanish and Some of Its Effects. Linguistic Inquiry 15. 103-130.

Travis, L. 1984. Parameters and Effects of Word Order Variation. Ph.D. Dissertation, MIT.

Vanelli, L., L. Renzi, P. Benincà. To appear. Typologie des pronoms sujets dans les langues romanes. Actes du XVII Congrès de Linguistique et Philologie Romane. Aix.

Wagner, R.-L. 1974. L'ancien français. Paris: Larousse.

Weinberg, A., I. Aoun, N. Hornstein, D. Lightfoot. 1986. Two types of locality. Ms. University of Maryland.

Wartburg, W. von. 1946. Evolution et structure de la langue française. Berne: Francke.

Zubizarreta, M.L. 1982. Theoretical Implications of Subject Extraction in Portuguese. The Linguistic Review 2.1.

Zwanenburg, W. 1979. Perte de la flexion nominale et fixation de l'ordre des mots en français medieval. XIV Congresso Internazionale di linguistica e filologia romanza. Naples.

OLD FRENCH EDITIONS CITED

La Chastelaine de Vergi, ed. by F. Whitehead. 1944. Manchester.

Eneas, ed. by J.-J. Salverda de Grave. 1925. Paris: Champion.

La Mort le Roi Artu, ed. by J. Frappier. 1964. Geneva: Droz.

Le Roman de Graal, ed. by B. Cerquiglini. 1981. Paris.

Robert de Clari. La Conquête de Constantinople. Historiens et Chroniqueurs du Moyen Age, ed. by A. Pauphilet. 1952. Gallimard.

Villehardouin, G. de. La Conquête de Constantinople. Historiens et Chroniqueurs du Moyen Age, ed. by A. Pauphilet. 1952. Gallimard.

James Blevins

Restructured Compounds and Romance Causatives

0 Introduction

This paper is concerned with some similarities between Romance causatives and the restructured Italian verbal complexes discussed in Rizzi (1978). It will be argued that such restructured compounds are neither the output of a syntactic movement rule nor of a rebracketing rule, but are instead base generated synthetic verbal compounds. Extending Zubizaretta's (1985) analysis of Romance causatives, it will be proposed that restructuring verbs are syntactically ambiguous between functioning as lexical heads and as bound morphemes, even though the phonology uniformly treats them as discrete words. This allows us to treat the 'restructuring' operation as a subcase of verbal compounding. Some of the peculiar properties of restructured complexes will be traced to their morphological derivation; in particular, their variable auxiliary selection and their status as non-composite constituents with respect to syntactic movement rules.

1 A Partial Description of Restructuring Verbs

The class of what Rizzi (1978) christened 'restructuring verbs' includes certain modal, aspectual and motion verbs. To begin with, let us focus on two of the distinctive characteristics of restructuring verbs: namely, auxiliary selection and clitic placement. What is striking about the auxiliary selection of restructuring verbs is that it is sensitive to the auxiliary selected by an embedded complement. In isolation, Italian modal and aspectual verbs occur with the auxiliary **avere** 'to have'. Thus, the past participle of **volere** 'to want' in (1) requires a form of **avere** as auxiliary.

(1) Mario **ha/*è** voluto un costoso regalo di Natale.
'Mario has wanted an expensive Christmas present.'

Yet, at Rizzi observes, when a verb which takes **essere** is embedded under a model or an aspectual, either auxiliary is possible. For instance, when **venire** 'to come', which invariably selects **essere**, is embedded under **volere** as in (3), either auxiliary is permitted.

(2) Piero **è/*ha** venuto con noi.
'Piero "is" come with us.'

(3) a. Piero **ha** voluto venire con noi.
b. Piero **è** volute venire con noi.
'Piero has/"is" wanted to come with us.'

Rizzi notices that the same class of verbs allow optional 'climbing' of clitic arguments of embedded infinitives. Clitics which are thematic arguments of verbs embedded under modal, aspectual or motion verbs are able to adjoin either to the embedded infinitives or to the matrix verb or its auxiliary. This optionality is illustrated in the sentences in (4). In (4a) the dative clitic argument of **parlare** 'to talk' attaches to that verb, while in (4b) the clitic climbs to adjoin to the matrix model verb **dovere** 'to have to'.

(4) a. Gianni deve parlargli personalmente.
'Gianni has to speak to him personally.'

b. Gianni **gli** deve parlare personalmente.
'Gianni to him has to speak personally.'

Rizzi suggests that these and other properties of restructured compounds are attributable to the fact that they can be associated with either of two distinct stuctural descriptions. To mediate between the pairs of available structures, Rizzi informally proposes a string-vacuous restructuring, or rebracketing rule of the sort given in (5).

(5) $[_{VP}\ V_r\ [_S (C)\ V_{inf}]]\quad [_{Pr}\ V_r\ (C)\ V_{inf}]$

Where V_r is a restructuring verb, C an optional complementizer, V_{inf} an infinitival verb and w a terminal substring of the form $[_{VP}\ V_r\ [_S\ (C)\ V_{inf}]]$, (5) optionally reanalyzes w as the complex terminal $[_{Pr}\ V_r\ (C)\ V_{inf}]$. **Pr** is here a complex predicate of somewhat indeterminate category. This rule is clearly not strongly structure-preserving, as it has the effect of deleting the null pronominal (PRO) subject that Rizzi assumes occurs in the embedded S. Nonetheless, a transformation of this sort does allow Rizzi to derive the exceptional behaviour of restructuring verbs with respect to cliticization. By way of illustration, consider the sentences in (6). The VP of (6a), which has not undergone restructuring, consists of a conjugated form of **dovere** and an infinitival complement. According to Rizzi's analysis, this complement consists of an infinitival S with a phonologically null subject. Hence, movement of the clitic **la** from its presumed postverbal base-generated position over the PRO subject gives rise to a Specified Subject Condition (SSC), or Condition A violation.

(6) a. Gianni [$_{VP}$ deve [$_{XP}$ presentarla a Fransesco]]
b. Gianni la [$_{VP}$ [$_{V*}$ deve presentare a Francecso]]
'Gianni had to present her to Francesco.'

In (6b), on the other hand, **deve** and the infinitival verb form a complex verbal predicate, in which the interfering subject has been 'reanalyzed' away. Accordingly, long movement of **la** does not run afoul of the SSC.

Rizzi argues that the interaction of clitic placement and auxiliary selection supports the view that the two exceptional properties of restructuring verbs are related to a single structural difference. On the assumption that constituents containing a restructuring verb and its putative complement have two possible structural analyses, he posits that both auxiliary change and clitic climbing are obligatory with restructured compounds and impossible with unrestructured constituents. That is, given two constructions containing restructuring verbs that minimally differ in that one occurs with **essere** and the other with **avere**, the postulated correlation predicts that noncomplement clitics will only be possible in the sentence containing the auxiliary **essere**. This prediction is borne out by the paradigm in (7).

(7) a. Mario ha dovuto venirci molte volte.
'Mario has had to come there many times.'

b. Mario c'è dovuto venire molte volte.
'Mario there has had to come many times.'

c. *?Mario ci ha dovuto venire molte volte.
'Mario there "is" had to come many times.'

In the unrestructured (7a), the locative clitic **ci** may, as expected, attach to the infinitival verb. As (7c) shows, **ci** may not, however, climb within the unrestructured sentence. Rather, climbing seems restricted to restructured compounds, in fact just to those where auxiliary change, if possible, has occurred.

1.1 Restructured and Causative Compounds

Rizzi identifies at least three respects in which restructured compounds differ from causative compounds. In the first place, matrix causatives never take the auxiliary of a verbal complement. Hence in (8), unlike (3b), **venire** cannot trigger the occurrence of **essere**, as **fare** invariably occurs with the auxiliary **avere**.

(8) Piero lo ha/*è fatto venire.
'Piero him has made come.'

(8) also shows that accusative clitics attach to matrix causatives, as was the case with restructuring verbs. Nonetheless, as Rizzi observes, causative compounds do not allow attachment of dative clitics associated with PP complements of an embedded verb. As we see in (9b), though, such cliticization is permissible with restructuring verbs.

(9) a. *?Mario gli farà scrivere Piero.
'Mario to him will have Peter write.'

b. Mario gli vuole scrivere.
'Mario to him wants to write.'

Further, while restructuring verbs freely take passive complements, causatives may only occur with complements in the active voice.

(10) a. Piero gli poteva essere presentato.
'Piero to him was allowed to be introduced.'

b. *Gianni ha fatto essere picchiato Piero da Mario.
'Gianni has had Piero be beaten by Mario.'

2 Burzio's Causative Rule

Before proposing a means of characterizing causative and restructuring compounds, let us briefly review Burzio's (1983) attempt to collapse the two types of compounds by means of a structure preserving syntactic operation. Burzio takes up the challenge of uniting these apparently discrete classes of verbs in terms of a general causative rule (CR). On his analysis, both restructured and causative compounds are the output of a VP-intraposition rule that adjoins the VP of an embedded S to that S. Thus, associated with sentence (11) is the underlying D-structure in (12a), which is converted via CR into the S-structure (12b).

(11) Giovanni fa leggere il libro a Mario.
'Giovanni made read the book Mario.'

(12) a. Giovanni fa [$_S$ Mario leggere il libro]
b. Giovanni fa [$_{VP}$ leggere il libro] [$_S$ a Mario _]

(The slash '/' simply marks the location from which the VP is extracted.) Burzio argues that the interaction of this analysis with Chomsky's (1981) Binding Theory and the general Government and Binding (GB) movement analysis of passives yields accurate predictions about the possible complements of causatives. In particular, he claims that these rules and principles correctly distinguish between such pairs as (13a) and (13b).

(13) a. Giovanni fa leggere il libro a Mario.
'Giovanni made read the book Mario.'
b. *Giovanni fa essere invitato (a) Mario.
'Giovanni made to be invited Mario.

The relevant difference is perspicuously displayed in the corresponding S-structures (14a) and (14b).

(14) a. Giovanni fa [leggere il libro] [a Mario].
b. *Giovanni fa [essere invitato e$_i$] [(a) Mario$_i$].

Burzio adopts the assumption, standard within GB, that the surface subjects of passive constructions are associated by the rule **Move-** with null anaphoric NP-traces ocurring in the object position of the passivized verb. Hence the passive complement in (14b), unlike the active embedded clause in (14a), contains an anaphoric trace which, in accordance with BT Condition A, must be bound by a local c-commanding antecedent. As a result, then, in (14b) but not (14a), VP intraposition produces a configuration in which an anaphor is locally free, in violation of Condition A.[2]

2.1 Restricting the Complements of Causatives

This result would constitute interesting support for the version of GB Theory that Burzio is defending, if not for the fact that it forces him to miss a relatively straightforward grammatico-thematic generalization. Elsewhere Burzio observes that **fare** must also be subcategorized for a VP, in order to accomodate examples like (15).

(15) Gianni fece [$_{VP}$ invitare Maria (da Piero)]
'Gianni has Maria invited (by Piero).'

He argues that the possibility of an oblique agentive **by**-phrase militates against interpreting the complement as an infinitival sentence with a null PRO subject, since there will be no available θ-role for the putative subject whenever the **by**-phrase occurs. Further, Burzio goes on to point out that (16) is ungrammatical even though it contains, on analogy to (15), a traceless VP.

(16) *Gianni fece [essere invitato (a) Maria (da Piero)]
'Gianni has be invited Maria by Piero.'

For theory-internal reasons, Burzio cannot appeal to Case Theory to rule out (16), since unaccusatives like **intervenire** in (17), which he also considers not to be Case-assigners, do occur with apparent complements embedded under **fare**.

(17) Gianni fece [intervenire Maria]
'Gianni has Maria intervene.'

Consequently, Burzio is driven to rule out examples such as

(16) on thematic grounds. The central difficulty with this solution is that by proposing completely different accounts for the ungrammaticality of (13b) and (16), Burzio misses the otherwise clear generalization that passives are universally impossible as complements of **fare**. Worse yet, the redundancy he introduces can be eliminated only at the expense of his proper binding account, given that such an account cannot in any obvious way be extended to cover cases like (16). On the other hand, any plausible semantic account of the illformedness of (16) is likely to be applicable in some form to cases like (13b).

Another redundant feature of Burzio's analysis is the dual subcategorization frames assigned to **fare**. The observation that **fare** must be subcategorized for a VP suggests a means of dispensing with this syntactic redundancy; namely, by positing that it uniformly subcategorizes for a VP complement. On this analysis, the VP of (11), repeated as (18a), would have the structure depicted in (18b).

(18) a. Giovanni fa leggere il libro a Mario.
b. $[_{VP}$ $[_{V'}$ fa $[_{VP}$ leggere il libro]] a Mario]

Thus, the VPs in (13) and (16) would be, in relevant respects, entirely parallel in structure. According to this way of looking at things, **fare** would then be a verb which combines with a VP to yield a lexical compound which takes the former external argument of the incorporated VP as an internal argument. When the complement VP is a monadic verb, the internal argument of the complex predicate shows up in direct object position. On the other hand, if the VP already consists of a transitive verb and direct object, the internalized external argument, for Case reasons, occurs as an oblique prepositional object.

2.2 More Evidence from Reflexives and Passives

However, there is reason to believe that this analysis, while feasible for French and Spanish causatives, is plainly incorrect for the Italian equivalents. As noted by Burzio (1981) and Zubizaretta (1985), Italian causatives do not allow complements containing the reflexive clitic **si**.

(19) a. On a fait **se** raser **Pierre**.
b. Lo hicimos **afeitarse** a **Pedro**.
'We made Peter shave himself.'

(20) a. On a fait **se** laver les mains à **Pierre**.
b. Lo hicimos **lavarse** las manos a **Pedro**.
'We made Peter wash his hands.'

(21) a. *Mario ha fatto **accusarsi Piero**.
'Mario made Piero accuse himself.'

b. *Mario ha fatto **lavarsi** le mani a **Piero**.
'Mario made Piero wash his hands.'
(From Zubizaretta (1985))

Whereas the French and Spanish sentences in (19) and (20) are fully wellformed, the similar Italian constructions in (21) are ungrammatical. This fact is unexplained and, in fact, altogether unexpected if **fare** takes a VP complement. On the other hand, if, as Zubizaretta suggests, **fare** combines with a verb to form a complex verbal predicate, this pattern is explicable, since a verb and **si** will form a constituent that does not satisfy the subcategorization frame of **fare**. Accordingly, (11) will be reassigned the structure in (22).

(22) $[_{VP}\ [_{V'}\ [_{V^\circ}$ fa leggere] il libro] a Mario]

To digress for a moment, it is clear that replacing (18b) with (22) has an additional consequence within theories of passivization such as those of Bach (1980) and Bresnan (1981). On the assumption that all transitive predicates passivize, the fact that $[_{V^\circ}$ fa leggere] is a complex ditransitive verb should lead us to expect that it, along with other causative constructions in Italian, will undergo passivization. Further, if, as suggested above, the French and Spanish correlates of **fare** take VP complements of the sort depicted in (18b), the direct object of the 'incorporated' VP should not be able to passivize. As (23)-(25) show, these expectations are borne out.

(23) a. Piero fece leggere quei brani (a/da Giovanni).
'Piero made Giovanni read those passages.'

b. Quei brani furono fatti leggere (a/da Giovanni).
'Those passages were made to be read by Giovanni.'

(24) a. On a fait construire la maison (à/par Casimiro).
b. Lo hicimos construir la casa (a/por Casimiro).
'We made Casimiro construct the house.'

(25) a. *La maison a été fait construire (à/par Casimiro).
b. *La casa fue hecha construir (a/por Casimiro).
'The house was made construct by Casimiro).
(From Zubizaretta (1985))

While (23b), the passive counterpart to (23a), is fully acceptable, the French and Spanish sentences in (25) are impossible.

Returning to the problem raised by Burzio's analysis, recall that we are faced with the task of excluding passive complements to **fare**. Since the solution presumably should extend to cover the equivalently bad constructions in French and Spanish, a syntactic account closely tied to the subcategorization differences noted above does not seem particularly promising. Let us explore then the possibility that the failure of causatives to accept passive complements has a thematic explanation. A solution with some intuitive appeal is provided by Zubizaretta, who claims that the thematic operations associated with causativization and passivation in Romance overlap in the sense that both deprive a predicate of its external argument. Consequently, the processes are incompatible as a subcase of a general prohibition against redundant operations or vacuous affixation. An analysis along these lines gains plausibility from the fact that it is not committed to some intrinsic incompatibility between causativization and passivization processes, a desirable result given that not all languages show the asymmetry that characterizes the Romance languages under discussion.

3 Argument Structure and Auxiliary Change

Let us now attempt to provide an account of restructuring compounds that correctly distinguishes them from the causative constructions just discussed. Of particular interest is the phenomenon of auxiliary variation which, though characteristic of restructuring compounds, never occurs with causative compounds. As a preliminary, let us review a key observation of Rizzi's. He notes that it is the final verb of a restructuring compound that wholly determines the choice of auxiliary for the compound. Thus, in (26a), the auxiliary is **avere**, since **prendere**, the last verb in the sequence, selects **avere**.

(26) a. Maria li avrebbe potuti stare per andare a
[+av] [+es] [+es]
'Maria them would have been able to be on the

prendere lei stessa.
[+av]
point of going to get herself.'

b. Maria ci sarebbe dovuta cominciare ad andare.
[+av] [+av] [+es]
'Maria there would "be" had to begin to go.'

Likewise, in (26b), **essere** shows up as the auxiliary of the compound, as its final verb, **andare**, selects **essere** in isolation.

Let us pursue the idea that restructuring verbs, like French and Spanish causatives, function as bound morphemes that combine with VPs to form V° predicates. Further, suppose that what distinguishes the two is that restructuring verbs, unlike causatives, preserve the selectional features and argument structure of the predicates they incorporate. Since the predicates are full VPs, all that remains to be preserved is their auxiliary selection and the agentivity of their external arguments. The examples in (27) should make clearer how a thematically neutral compounding process accounts for the observed pattern of auxiliary selection and clitic placement with restructured complexes. While the matrix VP of (26b) would have the bracketing represented in (27a), its unrestructured counterpart from (27b) would have the structure in (27c).

(27) a. $[_{VP}[_{V}$dovuta $[_{VP}[_{V}$cominciare $[_{VP}$ad andare]]]]]
b. Maria avrebbe dovuta cominciare ad andarci.
c. $[_{VP}$ dovuta $[_{XP}$ cominciare $[_{XP}$ ad andarci]]]

Initially **cominciare** combines with **ad andare** to form a V° with the selectional properties of **andare**. This V° subtends a VP which **dovuta** subsequently combines with to produce a V°, again with the original auxiliary selection features of **andare**. In the unrestructured (27c), however, **dovuta** alone is entered under the V° node, so that its intrinsic selectional properties fully determine the choice of auxiliary. This restructuring process is, as desired, string vacuous, and avoids assigning multiple syntactic subcategorization frames in favor of a unique subcategorization frame that yields either a lexical or phrasal unit.

3.1 Structural and Thematic Constraints on Clitic Interpretation

The assumption that restructuring is thematically neutral provides an account for the fact that the interpretation of clitics is not affected by the restructuring process. On the other hand, it is not unexpected that clitic construal should be influenced by causativization, given that such an operation is presumed to affect a verb's argument structure. Consequently, the contrast noted in (11) regarding the placement of dative clitics is, on this account, related to the thematic differences between causative and restructuring verbs. This is desirable, as there appears to be evidence that the traditional SSC account is untenable. First of all, as Manzini (1983) notes, the acceptability of a construction is directly related to its interpretation, so that the examples in (28) are acceptable if **gli** is the writer and unacceptable if **gli** is the recipient. An account that attributes the ungrammaticality of the first interpretation listed under (28a) to the fact that the clitic's anaphoric trace was locally free would presumably be committed to different structures for the two readings of the sentence.

(28) a. Gli fece scrivere (una lettera) a Maria.
*'To-him I-made Maria write (a letter).'
'Him I-made write (a letter) to Maria.'

b. L'uomo a cui fece scrivere (una lettera) a Maria.
*'The man to whom I-made Maria write (a letter).'
'The man whom I-made write to Maria.'
(From Manzini (1983))

Yet this explanation does not cover (28b) in any obvious way, as the gap in that sentence is a variable bound by a wh-operator and hence immune to Condition A. Furthermore, the postulated opacity of the VP does not affect lexical anaphors like **se stesso** which may be bound by a matrix subject, as in (29).

(29) **Piero** ha fatto lavare **se stesso** a/da Maria.
'Piero made Maria wash him(self).'

The intuitive core of a thematic account would be connected with the observation that a system of pronominal clitics which contains only an accusative and dative form cannot accomodate the three-way distinction required for the ditransitive causative compounds. Thus, as a means of resolving ambiguity, the two available clitics are associated with the internal and the internalized external arguments of the incorporated verb, to the exclusion of its original oblique arguments. However, let us now drop this somewhat tangential discussion and return to questions concerning the constituency of restructured compounds.

3.2 Constituency and Lexical Integrity

A clear prediction made by this account is that the putative complements of restructuring verbs will be expected to behave as syntactic constituents just in case restructuring (i.e., morphological combination) has not occurred. This conclusion receives a strong measure of support from the constituency tests (from Rizzi (1978)) given in (30)-(32).

(30) a. Mario sinceramente vorrebbe - ma a mio parere non potrà mai - **pagargli** interamente il suo debito.

'Mario sincerely would like - but in my opinion he will never be able - to pay him this debt.'

b. *Mario sinceramente **gli** vorrebbe - ma a mio parere non potrà mai - pagare interamente il suo debito.

'Mario sincerely him would like - but in my opinion he will never be able - to pay his debt.'

On the assumption that the climbing of **gli** is a reliable indication that restructuring has occurred, and that Right Node Raising (RNR) of the sort illustrated above applies exclusively to constituents, it follows that **pagargli interamente il suo debito**, but not **pagare interamente il suo debito** qualifies as a constituent. Likewise, while Complex NP Shift (CNPS) may shift a complex (i.e. long) postverbal complement to sentence final position in (31b), this movement cannot apply to (32a), which is, as the extracted clitic **ti** indicates, a restructured compound. Forced application results in the ungrammatical (32b); for our purposes it does not matter whether the NP **la mia idea** or the PP **a Firenze** is the moved element. In either case, the reasoning is that the string **ad esporre la mia idea** is not a syntactic constituent in (32).

(31) a. Fra qualche giorno, verrò ad esporti la mia idea a Firenze.

'In a few days, I will come to explain to you my idea in Florence.'

b. Fra qualche giorno, verrò a Firenze ad esporti la mia idea.

'In a few days, I will come to Florence to explain to you my idea.'

(32) a. Fra qualche giorno, ti verrò ad esporre la mia idea a Firenze.

'In a few days, to you I will come to explain my idea in Florence.'

b. *Fra qualche giorno, ti verrò a Firenze ad esporre la mia idea.

'In a few days, to you I will come to Florence to explain my idea.'

Clearly movement rules, like rules of clitic placement treat restructuring compounds more like unanalyzable words than like full clauses. Unlike simple lexical items, however, restructuring compounds allow the intrusion of a variety of types of adverbs. In (33), (again from Rizzi), the adverb **subito** occurs between the restructuring verb and its putative complement.[3]

(33) La verrò subito a scrivere.
'(I) it will come at once to write.'

The problem this introduces does not, however, arise uniquely in the case of restructured complexes, since the occurrence of such modifiers is similarly attested in nominal compounds. In Dutch sentences like (34a), from Sproat (1985), prepositions, articles and adjectives may freely occur in compounds inserted under an N° node.

(34) a. de [$_N$ van de boom vallende appels eter]
the from the tree falling apples eater
'the eater of apples falling from the tree'

b. de [$_N$ vlug van de boom vallende appels eter]
the quickly from the tree falling apples eater
'the eater of apples falling quickly from the tree'

Further, as (34b) shows, even adverbial modifiers are permitted in such compounds.[4] Yet (34b) is an uncontroversial compound. Thus, it would seem that if restructuring compounds cannot be neatly assimilated to a lexical paradigm, the fault lies more with an excessively restrictive paradigm than with the present account of restructuring compounds.

4 Conclusion

In summary, this analysis claims that Italian causative and restructured complexes are the output of compounding processes with distinct categorial inputs and thematic outputs. Both restructuring and causative verbs are treated, following essentially Zubizaretta, as species of bound morphemes, though restructuring verbs are also taken to be able to function as heads of clausal constituents. The general incompatibility of passives under causatives is attributed to the interaction of the thematic operations connected with causativization and passivization; an interaction that does not occur with the thematically neutral restructuring operation. It should be acknowledged that this account implicitly adopts a model of grammar which allows non-isomorphic syntactic and phonological representations and that it is assumed that these representations can be related by some version of the Marantz/Sproat mapping principle. From a current perspective, the central problem with Rizzi's transformational solution is that it must destroy structure introduced by the phrase structure component. In particular, the null subjects of infinitival clauses must be deleted, an operation that is inconsistent with any strong version of Chomsky's Projection Principle. Here, we avoid violating the Projection Principle at the cost of dispensing with its arguably gratuitous extension: the stipulation that verbal clauses uniformly require subjects.

NOTES

[1] I am indebted to Lee Baker, Denis Bouchard, Juliette Levin, Alec Marantz and Peter Sells for their helpful comments on earlier drafts. This work was supported in part by NSF Grant BNS-8318518 to the University of Texas at Austin.

[2] The displaced VP in both cases c-commands its former position, though this is of no particular importance as Burzio's causative rule makes use of a version of the projection principle according to which an intraposed VP does not leave a trace as it does not occupy an A-position at D-structure.

[3] These compounds also contain inflectional morphology, a fact whose potentially problematic character was pointed out to me by D. Bouchard (p.c.). However, as Vogel (1986) observes, inflectional morphology productively occurs within compounds in Italian. A few of her examples are repeated below:

a. lava piatti	'dish washer'	(piatti: 'dishes')
b. porta chiavi	'key chain'	(chiavi: 'keys')
c. apri bottiglie	'bottle opener'	(bottiglie:'bottles')

[4] Thanks to J. van der Wilt for this fact.

REFERENCES

Bach, E. (1980) "In Defense of Passive," **Linguistics and Philisophy** 3.3.

Bresnan, J. (1981) "The Passive in Lexical Theory," in **The Mental Representation of Grammatical Relations,** MIT Press, Cambridge, Massachusetts.

Burzio, L. (1983) "Conditions on Representation and Romance Syntax," **Linguistic Inquiry** 14.2, 193-221.

-(1981) **Intransitive Verbs and Italian Auxiliaries,** Doctoral dissertation, MIT, Cambridge, Massachusetts.

Chomsky, N. (1981) **Lectures on Government and Binding,** Dordrecht: Foris Publications.

Goodall, G. (1984) **Parallel Structures in Syntax,** Doctoral dissertation, UCSD, San Diego.

Manzini, M.R. (1983) **Restructuring and Reanalysis,** Doctoral dissertation, MIT, Cambridge, Massachusetts.

Marantz, A. (1984) "Tagalog Reduplication is Affixation, too," Paper presented at LSA Winter Meeting, Baltimore.

Rizzi, L. (1978) "A Restructuring Rule in Italian Syntax," in S.J. Keyser, ed., **Recent Transformational Studies in European Languages,** Linguistic Inquiry Monograph 3, MIT Press, Cambridge, Massachusetts.

Rouveret, A. and J.-R. Vergnaud (1980) "Specifying Reference to the Subject," **Linguistic Inquiry** 11, 97-202.

Sproat, R. (1985) **On Deriving the Lexicon,** Doctoral dissertation, MIT, Cambridge, Massachusetts.

Vogel, I. (1986) "Inflection before Compounding in Italian Lexical Phonology," Talk presented at Fifteenth Annual Milwaukee Morphology Meeting.

Zubizaretta, M.L. (1985) "The Relation between Morphophonology and Morphosyntax: The Case of Romance Causatives," **Linguistic Inquiry** 16, 247-289.

Denis Bouchard

En-Chain

In French, the clitic **en** can pronominalize a **de** X^n sequence. En can come from an object position (1), inside an object (2) or inside a subject (3).

(1) a. Je veux **de ce gâteau**.
'I want of that cake.'

b. J'en_i veux t_i.

(2) a. J'ai lu la préface **de ce livre**.
'I have read the preface of that book.'

b. J'en_i ai lu la préface t_i.

(3) [La préface t_i] en_i a été lue à la télé.
'The preface of-it has been read on the TV.'

The derivation of an EN-AVANT construction as in (3) (cf. Ruwet 1972, Milner 1978, Couquaux 1981, Haïk 1982) is now generally assumed to be similar to that of regular EN-constructions as in (1-2). The EN in all cases comes from inside a post-verbal NP, the difference only being that this NP has been subsequently preposed to the subject position in the case of EN-AVANT (cf. Couquaux 1981. Burzio 1981 has a similar analysis for Italian ne). Thus, we find EN-AVANT in constructions where the surface subject is a derived one like in passive (3), unaccusative verbs (4), predicate adjectives (5), raising verbs (6), but not if the subject is non-derived as in (7).

(4) a. L'idée **de partir** est venue à tous.
'The idea of to-leave is come to all.'
'The idea of leaving came to all.'

b. [L'idée t_i] en_i est venue à tous.

(5) a. La préface **de ce** **livre** est flatteuse.
'The preface of this book is flattering.'

b. [La préface t_i] en_i est flatteuse.

(6) a. [La préface t_i] en_i semble flatteuse.
'The preface of it seems flatterling.'

b. [La préface t_i] semble en_i être flatteuse.
'The preface seems of-it to be flattering.

(7) a. L'auteur **de ce** **livre** veut convaincre le
'The author of that book wants to convince the

lecteur de son impartialité.
reader of his impartiality.'

b. [L'auteur t_i] (*en_i) veut (*en_i)
convaincre le lecteur de son impartialité.

All the cases of EN-AVANT would fall under the schema in (8).

(8) NP_i en_k V ... $[_{NPi}...t_k...]$

Another quite broadly held assumption is that the choice of auxiliary, **être** or **avoir**, depends on whether or not a subject is derived: **être** is the AUX when the subject is derived (except for raising: cf. note 1), **avoir** when it is not (cf. Burzio 1981, Couquaux 1981). If we combine these assumptions about EN-AVANT coming only from derived subjects and **être** being the AUX selected for derived subjects, the prediction is that EN-AVANT should be possible if and only if **être** is the potential AUX.[1]

I will discuss three apparent problems for such an analyis of EN-AVANT.

Problem 1: As often discussed (cf. Kayne 1975, Milner 1978, Couquaux 1981, Haïk 1982, among others), EN-AVANT is only possible from inside a derived subject which is partitive (9a) or genitive (9b), not if it is either quantificational (10a) of indefinite plural (10b) (examples from Haïk 1981).

(9) a. Partitive:
Beaucoup en sort gâtées, de ces pommes.
'Many of them are spoiled, of these apples.'

b. Genitive:
La résumé en semble intéressant.
'The abstract of it seems interesting.'

(10) a. Quantitative:
*Beaucoup en sort gâtées, de pommes.

b. Indefinite plural:
*Des livres, en semblent intéressants.
'Of books, of-them seem interesting.'

But this restriction is not found when **en** comes from a non-moved object.

(11) a. Quantative:
Tu en as mangé beaucoup, de pommes.
'You of-them have eaten many, of apples.'

b. Indefinite plural:
J'en ai vu, des mauvais films.
'I of-them have seen, of bad movies.'

The problem here is that not all derived subjects allow EN-AVANT.

Problem 2: In reflexive constructions, the AUX is **être**. One way to account for this is to assume that (12a) is derived from (12b) by NP movement.

(12) a. Jean se rase.
'Jean himself shaves.'

b. △ se rase Jean.

We then expect EN-AVANT to be possible, but it is not.

(13) Quand l'auteur (de LSLT) s'en regarde dans
'When the author self of-it looks-at in

la glace le matin, à quoi pense-t-il?
the mirror in the morning, about what thinks he?'

'When the author of it looks at himself in the
mirror in the morning, what does he think about?'

There are some restrictions on clitic sequences in French, but this is not what rules out (13): some **se+en** sequences are acceptable, like when there is reflexivization of a dative complement (14a) or in frozen expressions (14b).

(14) a. Elle s'en achètera une paire, de
'She to-herself of-them will-buy a pair, of

souliers.
shoes.'

b. Il s'en va.
'He himself of-it goes.'
'He is leaving.'

(14a-b) are not cases of EN-AVANT, but even EN-AVANT is possible with **se** in middle and neuter constructions (cf. Ruwet 1972, Morin 1981).

(15) Middle:
Ce veston, la doublure s'en lave en dix
'This coat, the lining itself of-it washes in ten

minutes.
minutes.'

(16) Neuter:
Le bruit s'en est vite répandu.
'The rumor itself of-it is quickly spread.'

This second problem is that a construction with AUX **être**, i.e. reflexives, does not allow EN-AVANT.

Problem 3: The third problem is that some verbs with AUX **avoir** allow EN-AVANT, as Morin (1981) noted.

(17) [L'envie t_i] lui en$_i$ a pris/manqué/démangé.
'The longing to-him of-it has taken/missed/itched.'

Thus, we have three problematic constructions. In the first two cases, EN-AVANT is not possible although the assumptions on EN-AVANT and AUX selection predict it should be; in the third case, the assumptions predict that EN-AVANT is impossible, but it is fine. Problems 2 and 3 have received little attention. Extraction from QP subjects on the other hand has been discusses a lot, but there are problems with the analysis. I will take two recent analyses, Couquaux (1981) and Haïk (1982) to illustrate the kinds of problems that arise and then I will present a revised analysis of EN-AVANT out of QP subjects which partly extends to problems 2 and 3.

Couquaux (1981) assumes that what rules out sentences like (10a) is proper binding: en does not c-command its trace at LF, assuming the structure in (18).

(18) *°$_{S}$[$_{NPk}$ Beaucoup t$_{i}$][$_{VP}$ en$_{i}$-sont gâtées]]
de ces pommes.

A sentence like (5b) is grammatical, on the other hand, because he assumes a rule of scission that removes the PP from inside the NP prior to the raising of this NP and of cliticization of en, with the result that en can c-command its trace as we see in (20).

(19) Scission (Couquaux 1981):
[$_{NP}$ N^{n} - [$_{PP}$ en]] ==>[$_{NP}$ N^{n}] [$_{PP}$ en]

(20) [$_{NP}$ La préface] en$_{i}$ est flatteuse [$_{NPk}$ e]
[$_{PPi}$ e]

In the scission analysis, the reason for the ungrammaticality of (18) is due to the fact that it is stipulated that scission is not operative in quantificational NPs.

There are problems with this analysis. Crucially, scission must not leave a trace or we are back to a proper binding violation in (20). It is unclear what kind of formal operation it is then. If it is a movement transformation, it goes counter the Projection principle (Chomsky 1981) in not leaving a trace. If it is some kind of restructuring rule, there is an empirical problem in that en does not always correspond to the rightmost constituent in the NP (contrary to what Cinque (1980) claims): (21) is not any worse than (20) and forces scission to be a movement rule that leaves no trace.

(21) Les phrases à montée, $[_{NP_j}$ l'analyse t_i de
'Ruwet's analysis of raising sentences is very

Ruwet] en_i est t_j t élégante.
elegant.'

Moreover, Couquaux presents an argument against Kayne's (1975) **en**-extraposition rule: we never find full **de**-complements extraposed.

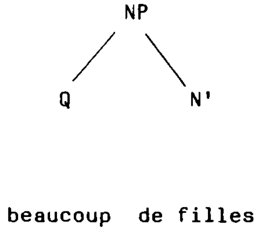
(22) *La préface est trop flatteuse **de ce livre**.

But exactly the same argument can be made against scission.[2]

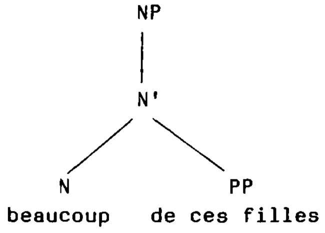
Another problem with Couquaux's analysis of (18) is that the contrast with respect to EN-AVANT between quantified NPs like (18) and genitives like (20) is purely stipulated: scission applies only to the latter.

Let us now turn to Haïk's analysis, where she tries to derive this stipulation. Haïk (1982) tries to solve this problem by using a structural difference between the two constructions that Milner (1978) argued for. The **de** X^n sequence in quantified NPs and indefinite plurals has head properties that the sequence in partitives and genitives does not have. Hence Milner assigns them different structures as in (23) and (24), respectively.[3]

(23) Q NPs, indef. plurals:



(24) Partitives, genitives:



This means that there is a crucial difference between cliticization of an **en** corresponding to **de filles** in (23) and cliticization of an **en** corresponding to **de ces filles** in (24): in a construction like (23), **en** is the head of the NP, but not in a construction like (24). Haïk assumes that for a NP to be interpretable, its head must be "visible" at LF. When the head is extracted, its trace must get Case to that effect. She assumes a genitivization rule of PRO in [Q PRO] when governed by V which changes PRO to **en**. So extraction from object position is fine. But in subject position, she assumes along the lines of Belletti and Rizzi (1981) that, although Case is assigned to the NP, it does not percolate to the head trace, hence the ungrammaticality.

There are problems with this analysis too. The non-percolation of Case to the head is surprising since Belletti and Rizzi use percolation as motivation for government of the head of a phrase by an external governor. Moreover, the **en** spellout of PRO involves much more than that. It is similar to Koopman's (1981) suggestion that **ce** be a spellout of PRO when it is governed in constructions like (25).

(25) Je sais ce que tu veux.
'I know that that you want.'
'I know what you want.'

The **en** spellout is subject to the same criticism as the **ce** spellout that Bouchard and Hirschbühler (1986) pointed out: it involves much more than spelling out and actually changes the category of the element from [+pro, +ana] to [+pro, -ana]. The proposed [Q PRO] structure also raises the problem of its distribution in infinitival clauses: why is (26) ungrammatical in that analysis?

(26) *[Beaucoup PRO partir tard] serait impoli.
'Many to-leave late would-be impolite.'

More importantly, the claim that Case is necessary for visibility at LF is also only a nonrevealing stipulation. True, such a claim about visibility of Case has been made elsewhere (Aoun 1981, Chomsky 1981), but is suffers from numerous conceptual and empirical problems that I have no time to go into here (but cf. Bouchard 1984: 43-51). Recall that the intuitive idea on which the visibility hypothesis is based is that features relevant to the rules of a component are visible at that level. But although person, number, gender are assumed to be relevant at PF, they are not sufficient to block contraction of **want-to** which can take place across PRO or non-Case-marked trace. Similarly, person, number, gender are relevant at LF but they are not sufficient to make an NP like **John** "visible" at LF: it must also have Case. So in fact, the visibility hypothesis is just the opposite of the intuition on which it is claimed to be based. More crucial to the discussion here is that it is not only Case that makes an NP visible at LF: PRO is said to be "inherently" visible since it can get a theta role although it has no Case (cf. Chomsky 1981: 334). But then the visibility account of the distribution of **en** collapses: if **en** is a spellout of PRO, then it should have its property of inherent visibility and not require Case assignment.

I will now present an alternative analysis that avoids these kinds of stipulations and derives the properties of EN-AVANT in Q NPs from independently motivated constraints on chain formation. If we assume some form of reconstruction for EN-AVANT, then the contrast between extraction of head **en** and non-head **en** can be explained. Essentially, the effect of reconstruction is that binding constraints apply as if material that has been displaced was still in its pre-movement position. So **en** in (5b) **la préface en est flatteuse** would properly bind its trace given reconstruction because [**la préface** t] is then considered to be in the c-command domain of **en**. However, I assume that reconstruction is not some operation that regurgitates derivationally old structure at LF. Rather, along the lines of Van Riemsdijk and Williams (1981), reconstruction is application of some process (like binding and conditions on it) at some intermediate stage before S-structure. I do not, however, assume that this is done at some specific intermediate level, like NP-structure in Van Riemsdijk and Williams. Rather, I assume as in Bouchard (1984) that the relation between an antecedent and an anaphor is defined by (27), not by free indexing and some condition on that indexing.

(27) ∂ AN-binds β if ∂ is the most local element that governs β and assigns its index to β.

This anaphoric relation holds between a clitic and the empty category that it binds. Reconstruction is simply allowing the null hypothesis where the antecedent can assign its index to the empty category at any point in the derivation and to consider that the anaphoric relation is then established regardless of what happens subsequently in the derivation. Crucially, the anaphoric relation is not subject to any further output constraint: it is determined by the indexing procedure instead. In the case of EN-AVANT, reconstruction in this sense has for effect that the en/empty category relation is established by (27) before raising of the NP into subject position as in (28).

(28) a. en_i est flatteuse $[_{NP}$ la préface $t_i]$ (AN-binding)

b. $[_{NPk}$ La préface $t_i]$ en_i est flatteuse t_k

This derivation is grammatical because both **en** and its trace and **la préface** and its trace have been properly linked by AN-binding, hence proper binding holds. So EN-AVANT is just as good as cliticization from object position because proper AN-binding can be established in both cases. Now consider the ungrammatical (29).

(29) *$[_{NPi}$ beaucoup $t_i]$ en_i sont gâtées e_i, de pommes.

It has the derivation in (30).

(30) a. en_i sont gâtées [beaucoup t_i] (AN-binding by en)

b. $[_{NPi}$ beaucoup $t_i]$ en_i sont gâtées e_i

Although **en** properly binds its trace t in (30a), the raised NP does not properly bind its trace e in (30b). The reason is that since **en** and the raised NP share the same index, **en** being the head of the Q NP as we saw in (23), the raised NP does not locally AN-bind e, **en** does. Hence the chain is broken and this results in ungrammaticality (cf. Rizzi 1982). So the contrast between (28) and (29) derives from the head/non-head property of **en** and its effects on representational chain formation.

Let us now turn to the **se** constructions. It is generally the case that EN-AVANT from Q NPs is not possible in alle the **se** constructions, be they reflexive/reciprocal (31), middle (32) or neuter (33).

(31) *Beaucoup s'en détestent, de gens.
'Many self-of-them-hate, of people.'

(32) *Beaucoup s'en sont vendus par milliers, de livres.
'Many self-of-them are sold by thousands, of books.'

(33) *Beaucoup s'en sont réunies, de gens.
'Many self-of-them are met, of people.'

This is expected given the representational chain analysis given above since **en** is head of the NP in all these cases and breaks the chain between the raised NP and the trace, assuming a deep structure roughly of the form in (A).

(A) △ se (AUX) V NP

On the other hand, if the subject is a derived one, we expect EN-AVANT from the genitive position to be possible. This is partly right: it is fairly acceptable in middles (34) and neuters (35).

(34) a. La doublure **de ce veston** se lave en 10 minutes.
'The lining of that coat washes in 10 minutes.'

b. La doublure s'en lave en 10 minutes.

(35) a. Le dénouement **de cette tragédie** se devine déjà.
'The end of this tragedy is already forseable.'

b. Le dénouement s'en devine déjà.

With reflexives/reciprocals however, EN-AVANT from genitive position is generally not acceptable: cf. (13) above and (36).

(36) a. Le président **de cette compagnie** se paie un
'The president of that company pays himself a

bon salaire.
good salary.'

b. ?*Le président s'en paie un bon salaire.

In my ear, however, (36b) is not as bad as (7) for example. Other examples can be constructed where reflexives/ reciprocals are not that bad, like (37).

(37) ??Beaucoup s'en détestent les uns les autres, de ces gens.

It is not clear to me why the facts are divergent in the EN-AVANT from genitive or partitive with reflexive **se**. If EN-AVANT from genitive was always possible, we could simply assume the analysis proposed above: since the AUX is **être** in all **se** constructions, the subject is a derived one. EN-AVANT from genitive is fine because **en** is not the head of the NP and no problem arises in the chain formation. If no EN-AVANT was possible with reflexive **se**, neither from Q NP nor genitive position, then we could assume, along the lines of Burzio (1981), that the reflexive **se** construction does not allow EN-AVANT because the subject is not a derived one, **se** cliticizing from object position. The split between middle/neuter and reflexive **se** could be due to a difference in derived versus nonderived status of the subject. The account of the AUX selection would then have to be slightly less straightforward.

But neither of these analyses account for the fact that some reflexive constructions seem to allow EN-AVANT from genitive position but others do not. It might be that one can maintain one of these analyses and dismiss the "delinquent" facts as falling under additional constraints. It has often been noted that **en** cliticization is subject to various constraints other than the ones discussed above and which do not seem to be easily amenable to general properties like headness versus nonheadness of **en**, for example. Thus Gross (1968) notes facts like the following (cf. also Morin 1981 on idiosyncracies of clitic constructions).

(38) a. Max a vu un grand lit/un certain lit.
b. Max en a vu un grand/*un certain.
c. Max en a vu certains.

(39) a. Luc a mangé tout un gâteau/tous les trois gâteaux.
b. Luc en a mangé tout un/*tous les trois.

(40) a. Il ne voit aucune solution.
b. Il voit chaque solution.
c. Il n'en voit aucune.
d. *Il en voit chaque.

(41) a. Le pied de la table/de Jean est énorme.
b. Le pied en est énorme (en=de la table/*de Jean).

(42) a. Il boit ce verre de ce vin.
b. Il en boit ce verre.
c. Le verre de ce vin est servi.
d. *Le verre en est servi.

Consider (41) and (42) for example. (41) shows that **en** is much better if it stands for an inanimate. (42) shows a case where a restriction holds, but only for EN-AVANT. Constraints of this sort are usually negative and exclude **en** from some contexts over and above more general conditions governing its distribution. Therefore, if what is going on with the reflexive construction is that additional conditions of this sort are excluding **en** in many cases, then the first solution suggested above should be preferred: the one where genitive **en** is predicted to be possible but does not always surface due to these additional constraints. Unfortunately, I cannot go much beyond this inconclusive state of affairs at this point on the topic of EN-AVANT in the reflexive construction.

Let us now turn to EN-AVANT in **avoir** AUX constructions (cf. Morin (1981)).

(43) a. L'envie lui en manque.
'He has no wish for it.'

b. Si l'envie t'en prend, fais-moi signe.
'If you feel like it, just tell me.'

c. L'envie m'en démangeait vraiment.
'I was really duying to do it.'

d. L'idée m'en plaît assez.
'I am interested by it.'

Other verbs that allow this are **sourire, transpirer, filtrer**. These verbs have three properties in common:

1° they are PSYCH-Vs, in the sense of Postal (1971) who claims that the constructions are derived by movement as in (44).

(44) NP_1	V	NP_2	==>	NP_2	V	NP_1
location		theme		theme		location

2° they enter in the impersonal construction:

(45) a. Il lui manque/prend l'envie de partir.
b. Il me démange vraiment/déplaît assez de partir maintenant.

3° they all have dative complements.

The second property is possibly derivable from the first one if we assume a movement analysis as in (44). As for the EN-AVANT problem, if the subjects are derived here, there is no problem in fact. But things are not that simple. Ruwet (1972) gives the following list of verbs as potential PSYCH-Vs.

(46) PSYCH-Vs: amuser, intéresser, agacer, ennuyer, effrayer, gêner, terrifier, horrifier, humilier, surprendre, étonner, impressioner, préoccuper, ...

All of these are quite bad with EN-AVANT.

(47) *La préface m'en intéresse/agace/...

Ruwet assumes that the verbs in (46) do not have derived subjects, contrary to Postal's proposal. One of the reasons for which Postal postulated a transformational rule here was to explain the switch in selectional restrictions and thematic roles as sketched in (44) above. But Ruwet argues that this oculd just as well be dealt with in the base component and that, in fact, this alternative is preferable because there are all sorts of irregularities in these mappings.

Note that a rule that allows NPs to switch positions as in (44) is not possible in current frameworks given conditions like the theta criterion and the projection principle. A possible derivation however is one as in (48).

(48) $\triangle$ V NP_1 NP_2
theme location

This is not possible in French, however, where the Phrase Structure component does not allow double object constructions. So the verbs in (46) do have to be base generated with a subject in French, as Ruwet claimed. This explains why EN-AVANT is not possible with them: they do not have a derived subject. But what about the verbs in (43)? They could be inserted in a base like (50) since this is a possible French deep structure.

(49) △ V NP_1 à NP_2

So maybe both Ruwet and Postal were right: some PSYCH-Vs, the accusative ones, are base generated with deep structure subjects; others, the dative ones, have a derived subject: the latter would be like dative unaccusative verbs. This accounts for the distribution of EN-AVANT. It opens another big can of worms which I have only space to peek into here, namely the problem of AUX selection in French. Why is **avoir** the AUX for dative PSYCH-Vs if they have derived subjects? Here are preliminary remarks.

Etre is the AUX that obligatorily blocks theta role assignment to the subject. So if the verb is usually a subject theta assigner, it must have **être** in derived subject constructions to avoid two theta roles from being assigned to the same element, like in passive, **se**-constructions.

Avoir is an AUX that can transmit a theta role. So it will be the AUX in a construction where a theta role must be transmitted to the subject, as in **Jean a fini son travail,** or when a theta role cannot be assigned since **être** could not block assignment of a theta role as it is required to do: this is the case of raising verbs (50a) and of **être** itself (50b).

(50) a. Jean a semblé vouloir en finir.
b. Jean a été blessé.

It would also be the case of dative PSYCH-Vs if they have the deep structure given in (49).

A problem immediately arises: the structure in (49) is similar to the one assumed for unaccusative verbs, but **être** is the AUX for these verbs, not **avoir**.

(51) Jean est/*a arrivé.

A difference between **être** taking verbs like passive verbs and ergative verbs on the one hand, and **avoir** taking verbs like raising verbs is that the derived subject in the former comes from the object position of the V, but not so for raising Vs. We would try to assimilate PSYCH-Vs with raising Vs by suggesting that they have a small clause structure as in (52).

(52) △ manque$[_{SC}$ l'envie de partie à Jean]

But it is not at all sure for me that this is the right direction to look into. For one thing, the intuitions on theta role assignment in this construction do not correspond to a small clause analysis. Rather, it seems that conditions on the AUX selection itself are needed.

In conclusion, although we have a simple solution to the first problem on EN-AVANT from quantified NPs, the representational analysis of the chain, the problems that arise for EN-AVANT in reflexive constructions and PSYCH-V constructions have been shown to be more complex than they appear and are related to another highly complex problem, the selection of AUX in French.

NOTES

[1] EN-AVANT is possible in constructions which take **avoir** as AUX, like the raising construction where the subject is also derived (cf. (6a)). This is possible only if there is no embedded V on which **en** can cliticize. Thus, compare (i) with (6b).

(i) *La préface en semble être flatteuse.

I will assume that the use of **avoir** in raising constructions is similar to its use in **être** constructions like passive (3), where **avoir** is the AUX of **être**, and I will restrict the discussion to "pure" **être** and **avoir** constructions.

[2] There are a few cases where we seem to get **de**-complement extraposition (of scission), but then **en** cliticization is not possible.

(i) a. Jean est fier.
'Jean is proud.'
b. Jean l'est.
'Jean it-is.'
'Jean is it (proud).'
c. Jean est fier de son travail.
'Jean is proud of his work.'
d. ?Jean l'est de son travail.
e. *Jean l'en est.

The ungrammaticality of (e) is not due to a constraint on clitic sequence as in (ii) since (iii) is acceptable.

(ii) *Jean me vous a présentés.
'Jean to-me you has introduced.'
'Jean introduced you to me.'

(iii) Marie l'en a dissuadé.
'Marie him of-it has dissuaded.'
'Marie dissuaded him of it.'

[3] Note that this difference in structure removes the need for the stipulation that scission does not apply in Q NPs in Couquaux's analysis: the **en** does not replace a PP in this case but a projection of N.

REFERENCES

Aoun, Y. 1981. The Formal Nature of Anaphoric Relations, doctoral dissertation, MIT.

Belletti, A. and L. Rizzi. 1981. The syntax of 'ne': some theoretical implications. The Linguistic Review, 1.2.

Bouchard, D. 1984. On the Content of Empty Categories. Dordrecht: Foris Publications.

Bouchard, D. and P. Hirschbühler. 1986. French **Quoi** and its clitic allomorph **Que**. Proceedings of the 15th Linguistic Symposium on Romance Languages, C. Neidle and N. Cedeno, eds. Dordrecht: Foris Publications.

Burzio, L. 1981. Intransitive Verbs and Italian Auxiliaries. Doctoral dissertation, MIT.

Chomsky, N. 1981. Lectures on Government and Binding. Dordrecht: Foris Publications.

Cinque, G. 1980. On Extraction from NP in Italian. Journal of Italian Linguistics, 5.1/2.

Couquaux, D. 1981. French Predication and the theory of Grammer, in J. Koster and R. May, eds., Levels of Syntactic Representation. Dordrecht: Foris Publications.

Gross, M. 1968. Grammaire transformationnele du français: Syntaxe du verbe. Paris: Larousse.

Haïk, I. 1982. On clitic **en** in French. Journal of Linguistic Research, 2.1.

Kayne, R. 1975. French Syntax: The Transformational Cycle. Cambridge: MIT Press.

Milner, J.-C. 1978. De la syntaxe à l'interprétation: quantités, insultes, exclamations Paris: Editions du Seuil.

Morin, Y.-C. 1981. Some myths about pronominal clitics in French. Linguistic Analysis, 8.2.

Postal, P. 1971. Cross-Over Phenomena. New York: Holt, Rinehart and Winston.

Riemsdijk, H. van and E. Williams. 1981. NP-Structure. The Linguistic Review, 1.1.

Rizzi, L. 1982. On Chain Formation. Università della Calabria. Paper presented at GLOW, York.

Ruwet, N. 1972. Théorie syntaxique et syntaxe du français. Paris: Seuil.

Becky Brown

Problems with Past Participles Agreement in French and Italian Dialects

0. Introduction

The topic of past participle agreement, though widely discussed in both traditional and current linguistic literature, remains for the most part a perplexing issue.[1] The apparent variable nature of the system in Romance languages poses a challenge for defining its properties. Generally, there seems to be disagreement on the exact nature of the triggering element in part participle (PP) accord. Traditional theory with respect to French and Italian holds that agreement occurs in the presence of certain lexical items. Thus in standard French, for example, PP agreement occurs in compound tenses having the auxiliary "to be", (1a), or an accusative clitic pronoun, (1b).

(1) a. Elle$_i$ est morte$_i$ ce matin.
'She died this morning.'

b. Il l$_i$'a faite$_i$ aujourd'hui.
'He made it (fem.) today.'

Current theory includes a variety of perspectives. Traces (Burzio 1981) movement (Kayne 1984), or coindexing (Bouchard 1985) can each trigger PP agreement.

Regarding Romance dialect data, traditional theory as well as some of the current theories fall short in their predictions. In this paper I will discuss the strengths and weaknesses of these theories in an effort to show that coindexing (Bouchard 1985) best explains the complex PP agreement system in Romance dialects. Examples will be presented from the following languages: Catalan, Provencal, Cajun and Cremonese. A few samples will also be taken from Corso di Bastía, Lazio, Occitan and Walloon. These languages all seem to have unique agreement systems that would rigorously test any given theory.[2]

1.0 Traditional accounts

Considering PP agreement from the traditional standpoint seems, for the most part, inadequate. Even though the grammar book rules account for a large portion of the data, they are grossly over-generalized.[3] And even still, they fail to capture generalizations in the agreement system. There are five basic processes involving PP agreement, none of which are related in traditional theory. They are: passivization, wh-movement, cliticization, ergative verb raising and reflexivization.[4] It will become evident later that due to the idiosyncratic and asymmetric nature of the dialect data, relating all of these processes is, in effect, a theoretical ideal.

Let us now consider some of the cases where standard theory falls short with respect to the dialects.

One of the principles rules of PP agreement in French and Italian is that in the compound tenses with the auxiliary "to be" (être/essere), the PP agrees with the subject.

(2) a. Il$_i$ est mort$_i$. Elle$_i$ est morte$_i$.

b. $_i$ E morto$_i$. $_i$ E morta$_i$.
'He died. She died.'

If the auxiliary is "to have" (avoir/avere), there is no agreement either with the subject or the object.

(3) a. Elle$_i$ a construit$_j$ la maison$_k$.

b. $_i$ Ha costrutto$_j$ la casa$_k$.[5]

'She built the house.'

On the contrary, in the Italian dialect in Lazio, the auxiliary "to be" does not trigger subject agreement.[6]

(4) a. $_i$ semo cercato$_j$
'we have looked'

b. $_i$ sete visto$_j$
'you have seen'

c. $_i$ semo vennuto$_j$
'we have sold'

And in the Catalan dialects of Balear and Valenciano (Badia Margarit 1962:227), there is object agreement with the auxiliary "to have".

(5) a. He vista$_i$ la mare$_i$.
'I have seen the mother.'

b. He trobats$_i$ els amics$_i$.
'I have met the friends.'

Additionally, in mainstream Catalan (Yates 1975:89) when the auxiliary is "tenir" used in the sense of "to have", the PP agrees with its object.

(6) Tinc pensada$_i$ una contesta$_i$.
'I have thought about the answer.'

There is, however, one instance in standard Italian where a PP agrees with its object. Belletti 1983 points out this particular structure that is cited in (7).

(7) Conosciuta$_i$ Maria$_i$, Gianni è cambiato del tutto.
'Known Maria, Gianni changed completely.'

This same structure also occurs in the Italian dialect of Cremonese.[7] Only here, there is no agreement.

(8) a. Rivāt$_i$ le me surẹle$_j$,...
'Arrived my sisters,...'

b. Sistemāt$_i$ la fačẹ̄nda$_j$,...
'Arranged the matters,...'

c. Pagāt$_i$ le spẹ̄ze$_j$,...
'Paid the purchases,...'

Another standard rule claims that PPs agree with preceding accusative clitic pronouns.

(9) a. Il l$_i$'a écrite$_i$. Il l$_i$'a écrit$_i$.

b. L$_i$'ha scritta$_i$. L$_i$'ha scritto$_i$.

'He wrote it (fem.). He wrote it (mas.).'

This rule is extended in French, but not in Italian, to preceding uncliticized direct objects, in other words, "wh-moved".

(10) a. la maison$_i$ que j'ai construite$_i$

b. *la casa$_i$ che ho construtta$_i$

'the house that I built'

In some dialects, though, these rules do not hold. In Cajun, Provencal and Walloon, for example, PPs do not agree with wh-moved nor cliticized direct objects.[8]

(11) Cajun: la cravate$_i$ que t'as fait$_j$
'the tie that you made'

Provencal: la terro$_i$ qu'ai croumpa$_j$
'the land that I bought'

Walloon: dju l$_i$'a vèyou$_j$
'I saw her'

A fourth deviation from the standard occurs in the dialect of Corso di Bastía.[9] Here there is accord with dative clitics.

(12) a. l$_i$'avia racummandati$_i$ di vruscill

(S.I.)'gli avavo raccomandato di bruciarla'
'I had recommended to them to burn it (fem.)'

b. li$_i$ l'avia cunsigliati$_i$

(S.I.)'glie l'aveva consigliato'
'S/he had advised it to him/them.'

c. li$_i$ l'ghui prumessi$_i$

(S.I.)'glie l'ho promesso'
'I promised it to him.'

From these counterexamples, we can surmise that because of the asymmetric nature of PP agreement, postulating general structural rules of accord is inadequate. Instead we would favor a more abstract approach aimed at underlying universals of agreement. Several current frameworks do, in fact, take this perspective. In the next three sections we will examine these theories.

1.1 Burzio 1981

Burzio claims that PPs agree with the subject position, the same position that triggers verb agreement. His theory predicts that a PP assumes the agreement features of a preceding NP whose original position (trace) is governed by the PP. These NPs are base-generated post-verbally. This category encompasses, therefore, direct objects and what he calls "i (inverted) - subjects". I-subjects occur in sentences such as "Gianni è venuto." ('Gianni came.') whose deep structure is (13).

(13) è - venuto - Gianni

After movement or "ergative verb raising", we get (14).

(14) Gianni - è - venuto - t

Now the conditions are met for Burzio's agreement rule: a PP agrees in gender and number with an element binding its direct object. This underlying structural contrast is meant to explain the discrepancy in accord in (15a) and (15b), where (15a) has the i-subject.

(15) a. Si è arrivati.
'One has arrived.'

b. Si è telefonato a Gianni.
'One has called Gianni.'

An advantage to this approach is that two traditionally separate agreement properties are related: cliticization and ergative verb raising. Furthermore, this theory predicts auxiliary selection between "to have" or "to be". He states that the auxiliary will be "to be" when a binding relation exists between the subject and a nominal constituent (direct objects and clitics) of the predicate. This is a marked advancement, since traditional theory makes no claim pertaining to this problem.

However, we encounter several problems with this framework in the Italian dialect of Cremonese due to its asymmetric agreement system. In ergative verb raising constructions, the PP agrees with the subject when the verb is in the first or second person. Oddly, there is agreement in the third person only when there is a redundant pronoun.

(16) a. La rǫba$_i$ ę́ rivāt$_j$.
'The stuff arrived.'

b. La rǫba$_i$ l'ę́ rivāda$_i$.
'The stuff, it arrived.'

In both (16a) and (16b), Burzio's conditions for agreement are met.

Complications arise also in cliticization. PPs agree with accusative clitics, (17) and (18), but not with "ne/na/en", (19).

(17) G'ǭ víst$_i$ to zía$_j$.
'I saw your aunt.'

L$_i$'ǭ vīsta$_i$.
'I saw her.'

(18) Kredi de vīla$_i$ vīsta$_i$.
'I believe to have seen her.'

(19) 'N$_i$ ǭ víst$_j$ tānta$_i$.
'I saw so much of it (fem.)'

'N$_i$ ǭ víst$_j$ tānte$_i$.
'I saw so many of them (fem.)'

According to Burzio, the binding relationship between the clitics and their traces would trigger agreement in all cases as in standard Italian.

(20) a. L$_i$'ho vista$_i$.
'I saw her.'

b. Gianni ne$_i$ ha ripresa$_i$ una$_i$.
'Gianni took one (fem.) of them.'

Another example of person asymmetry can be found in cliticization in Catalan. Agreement occurs only with third person accusative clitics, including "en/ne", (21c).

(21) a. He ecrit la invitació però encara no l$_i$'he tramesa$_i$
'I have written the invitation but I haven't sent it yet.'

b. Heu entès les preguntes? --No, no les$_i$ hem enteses$_i$
'Did you understand the questions? --No, we didn't understand them.'

c. Són boníssimes aquestes prunes; jo n$_i$'he menjada$_i$ una$_i$ i aguest se n$_i$'a menjades$_i$ dues$_i$ o tres$_i$.
'These plums are very good; I have eaten one and he has eaten two or three of them.'

Another aspect of Burzio's theory, namely auxiliary selection, while being a strength in terms of standard Italian and French, loses its predictive power in some of the dialects. In Cajun, for example, the selection is variable. In some Cajun communities "to have" is the only auxiliary used in the compound tenses, and there is no agreement. Here, Burzio's analysis holds. The majority of Cajun speakers, however, choose between the two auxiliaries, but the selection is different from standard French. Verbs that take "to be" in standard French sometimes vary between "to be" and "to have" in Cajun.

(22) a. Ils ont venu. / Ils sont venus.
'They came.'

b. J'ai né. / Je suis né.
'I was born.'

PPs of verbs conjugated with "to be" appear to agree in gender and number with the subjects.

(23) a. Il$_i$ est mort$_i$.
'He died.'

b. Sa maman$_i$ est morte$_i$.
'His mother died.'

According to Burzio, then, what are usually i-subject structures in standard French, are not in Cajun. The set of ergative verbs in Cajun is a subset of those in standard French. Apparently, the variable choice of auxiliary in (22) would be difficult for any theory to account for.

Finally, Burzio would have trouble explaining auxiliary selection in Lazio. There is asymmetry depending on person. First and second person take "to be", while third takes "to have". And there is no agreement in either case.

(24) $_i$ só mañate$_j$ $_i$ seme mañate$_j$

$_i$ si mañate$_j$ $_i$ séte mañate$_j$

$_i$ a mañate$_j$ $_i$ au mañate$_j$

'I ate, etc.'

In summary, Burzio's theory cannot cover the cases where there is any kind of person asymmetry, (cf. ergative verbs in Cremonese, cliticization in Catalan, and auxiliary selection in Lazio). We also saw that in some of the dialects ergative constructions do not necessarily entail agreement (cf. (16a)), while some non-ergative constructions do, (cf. (5) and (6)).

Now let us consider agreement in terms of small clauses as proposed by Kayne.

1.2 Kayne 1984

According to Kayne, agreement is a local phenomenon and should parallel subject-verb agreement. His idea is to break up the Aux-PP constituent by positing that a PP can have its own subject.[10]

(25) Jean les$_i$ a ((e)$_i$ repeintes (e)$_i$)
'Jean repainted them (fem.).'

Concerning agreement, the PP agrees with its own subject. This small clause analysis, he claims, is preferable to

Burzio's because the lack of agreement in (26a) illustrates that agreement is not with the trace in object position in (26b).

(26) a. Jean a repeint$_i$ la table$_j$.
'Jean repainted the table.'

b. Jean l$_i$'a repeinte$_i$ t$_j$.
'Jean repainted it (fem.).'

Moreover, he feels his theory is stronger in that it accounts for several problematic cases for previous analyses where the direct object has been moved in front of the PP, but yet agreement is possible.

Evidence, though, in Catalan suggests that Kayne's structure is not necessarily better. The Aux-PP constituent cannot be broken up as Kayne would have it. Yates 1975 writes in the notes on the compound tenses, "The auxiliary cannot be separated from the PP, not even by a subject pronoun." (p.89)

(27) a. Vostè ha viscut sempre aqui?
b. Sempre ha viscut aqui vostè?
c. *Ha vostè viscut sempre aqui?
d. *Vostè ha sempre viscut aqui?
'Have you always lived here?'

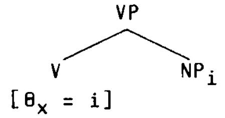
The main objection to this analysis is the premise upon which it rests: agreement is a local phenomenon. We know than it CAN be, but generally it is not. Moreover, Bresnan and Mchombo 1985 discuss the similarities between grammatical and anaphoric agreement within the grammatical structures of one language. A noteworthy point is that person, number and gender are the same pronominal categories which universally show agreement in anaphora. With this idea in mind, we should perhaps be dealing with agreement in a more general sense.

A second objection to a small clause analysis is that it creates structure, and a complex one at that. We will see later that we can exploit a mechanism already incorporated in the theory to explain agreement. Also, it can deal with local and non-local agreement. Most significant with respect to the dialects, though, is that this theory, like Burzio's, does not account for agreement systems with person asymmetry. Consider now a third approach: Bouchard's V-index analysis.

1.3 Bouchard 1985

Bouchard presents a framework where agreement, whether local or not, is always triggered by coindexing. He claims that PP agreement is triggered by coindexation between the PP and its direct object, the direct object entering its index in the theta-grid of the verb to get its theta-role. The idea originates in Stowell 1981 where it is suggested that a formal way to express theta-assignment from a verb to an argument is to say that the index of the complement is entered in the theta-grid of the verb. Consequently, the verb and the complement are coindexed. A simple illustration of the mechanism is in (28).

(28)



Moreover, Bouchard assumes that there are two parts of an index, a syntactic (s) and referential (ref) index. The S-indices are for coindexing elements that are not referential to indicate a syntactic relation. These indices are found, for example, in determiner-adjective constructions to account for their agreement with a head NP (la vieille dame). The ref-indices are, as expected, for coindexing referential elements.

Crucial to the approach is Condition C of the Binding Theory (which states that R-expressions must be free), and the ordering of processes in the grammar. Bouchard argues for the following (simplied for the purposes of this paper):

(29) Agreement > Binding Theory > V-Ref-Indexing

Since Condition C is a principle governing referential dependencies, only ref-indices are relevant for the binding theory. The implication is that in (30), V-Ref-Indexing would induce Binding Theory violations since "Marie" would not be free. Hence, V-ref-indexing can only take place after the Binding Conditions apply in the grammar.

(30) Jean a vu_i $Marie_i$.

Through several examples, then, he illustrates that V-Ref-Indexing must occur later than the Binding Theory so that R-expressions remain free, and thereby, avoid violating Condition C.[11] He goes on to state that a priori nothing forces this ordering between indexing and agreement. For example, in Old Italian and Old French, since PP and object agreed, V-Indexing took place before agreement.

Bouchard's idea is attractive for dialect agreement for several reasons. Consider the Romance agreement properties in table form. Then through discussion, the reasons will become apparent.

(31)	Pass.	Erga-tives	Post-V. NPs	Wh-Moved	Acc. Clitics
Provencal	+	+	-	-	-
Cajun	+	+	-	-	-
Cremonese	+	±*	-	-	+
French	+	+	-	+	+
Italian	+	+	±	-	±*
Catalan	+	-	+	+	±*
Walloon	#	-	-	#	-
Lazio	#	-	#	#	#

(where "*" =person asymmetry and "#" =data unavailable)

First, we note that passivization is the unmarked case across the board. For this reason, we may be prematurely attracted to Kayne's idea of treating his small clauses as passives. Next, with the exception of Lazio, Walloon and Catalan, ergative verbs generally entail agreement. Thus, Burzio's framework looks promising since he relates the two processes of auxiliary selection and agreement. Neither of the theories, however, accounts for the variation in the other problem areas: wh-movement, cliticization and post-verbal NPs. We need a theory that will not treat clitics uniformly since we know that in a given language there is PP agreement with some clitics, but not with others. Moreover, there is person asymmetry. Similarly, we are at fault if we speak in terms of the post PP position as having general properties when we noted that agreement is variable. And thirdly, we saw that movement does not necessarily trigger agreement. What we conclude, then, is that not only must we abandon a categoric approach (the traditional view), but also we cannot rely on underlying structural distinctions in the sense of Burzio and Kayne. Rather, the presence of absence of agreement can be viewed in terms of two types of indices within the theta-grid of the PP where one triggers agreement (V-S-indices), and the other does not (V-Ref-Indices).

If we discuss PP agreement in these terms it seems that perhaps we can account for several of the problem cases mentioned earlier. For example, we have seen in constructions in Catalan with the auxiliary "tenir", there is agreement with a following direct object (6). According to Bouchard, then, we would have to take in consideration the Binding Conditions and the type of V-indexing. The PP with "tenir" would have to have V-S-indexing which would trigger agreement while at the same time avoid violating the Binding Conditions. That is, in (32) the V-S-indexing feeds agreement, but is not relevant to the Binding Conditions. Hence, "contesta" remains "free" in its governing category.

(32) Tinc pensada$_i$ una contesta$_i$.
'I have thought about the answer.'

Then we saw that in Lazio the auxiliary "to be" does not entail agreement (4). The parameter must be V-ref-indexing with a bleeding effect on agreement.

This approach also looks desirable for Cajun where the choice of the auxiliary for one verb is variable. If the speakers opts for "to have", there is never any agreement, even with movement. If "to be" is chosen, however, there will be agreement. The "to be" cases, then, would carry S-indices that trigger agreement.

In Cremonese where there is agreement asymmetry depending on person, we would claim that in this dialict first and second person carry S-indices to feed agreement, while third does not.

Theoretically, we could proceed through each of the dialects explaining different phenomena in terms of two types of indices.

2.0 Conclusion

In this paper, I have illustrated the complex, diverse, and asymmetric nature of the PP agreement system in various Romance dialects. The ideal theory would attempt to account for all of the idiosyncratic behavior. Some current theories do extend to a certain degree to several dialects. Perhaps only slight modification would be necessary. On the other hand, other dialects remain unexplained, and certain structures, especially cliticization and post-verbal NPs, cause problems with the most rigorous theories.

From the nature of the data presented, it seems apparent that we would favor a model that could be flexible enough to handle a great deal of asymmetry and variation. For example, we might not necessarily want the theory to predict certain phenomena regardless of person. As it turns out the cases of person asymmetry are so numerous, that they themselves merit further investigation. Additional evidence comes from the Gerona dialect of Catalan in the region of La Garrotxa where "to have" is the usual auxiliary. However, "to be" occurs as the auxiliary in determining verb person, notably first.

(33) soc anat, soc begut[12]
'I went, I drank'

Person asymmetry is even noted in standard Italian where first and second person cliticization agreement is optional, but obligatory in third.

(34) a. Ragazza, chi vi ha vista/o?
'Girl, who saw you?'

b. Chi la ha vista/*o?
'Who saw it (fem.)?'

The theory would also have to handle variation in other clitic behavior. The exceptional cases we have observed are dative agreement in Corso di Bastía and no agreement with "ne" in Cremonese and French. The data indicate that clitics do not behave as a class with respect to agreement.

Finally, we would also like the theory to be able to extend to the oddest cases of agreement. In (35), note that Occitan is characteristic of agreement on the second auxiliary!

(35) Aquesta musica, l'avetz aguda/*agut escotada.
'This music, you have had heard it.'

For these reasons, it appears that exploiting the indexing in the theta-grid is an economical approach that can extend to asymmetric phenomena.

NOTES

[1] I am grateful to Richard Kayne and Denis Bouchard for providing insightful comments on earlier drafts and for patiently discussing many aspects of this study. I am equally appreciative of Antonella Pease, Fiona Mc Laughlin, Leo Villalón, Jean-Pierre Montreuil, Carolyn Mackay, and Lourdes Mir for their help with the data.

[2] In this paper I use the terms "language" and "dialect" interchangeably. I am intentionally avoiding the issue on the classification of the data in this study. Of importance for now is that these are Romance data, and therefore, should have some unifying properties.

[3] This has been pointed out in Burzio 1981 and Kayne 1984.

[4] Reflexive agreement is a complex and involved topic in itself, and therefore will not be discussed in this paper. Some of the agreement problems one encounters are accusative clitic interference in standard Italian, post-verbal NP interference in Provencal, and "to have" as the auxiliary in Cajun, Veneto, and Toscano. We will leave these cases for further research when agreement in general is more readily understood.

[5] There is agreement here between PP and object in formal style and Old Italian. We will deal with this later.

[6] The data on Lazio is from Rohlfs 1969.

[7] Cremonese is well documented in Tossini 1975.

[8] The Cajun data throughout this paper are from Conwell and Juilland 1963. The examples in Provencal are from Fourvières 1941. The Walloon cases, taken from Remacle 1956, are especially interesting since there exist feminin forms

in the grammar (i.e. vèyou/vèyowe, trompé/trompêye), yet they do not appear in the PP forms.

[9] Rohlfs 1969 provides the data on Corso di Bastía.

[10] Regarding case assignment, the EC subject-PP structure behaves as passives do. That is, accusative case is assigned by the PP (not by the auxiliary) to its object position.

[11] For more discussion, see Bouchard pp. 6-10.

[12] Badia Margarit 1962:293. My informant adds that "to be" is the auxiliary as well in the first person plural, and there is no agreement.

REFERENCES

Badia Margarit, A. 1962. Gramatica Catalana. Madrid: Editorial Gredos.

Bailon, C. and R. Lafont. 1969. Metode per aprene l'occitan parlat. Montpellier: Centre d'Estudis Occitans.

Belletti, A. 1983. "Frasi ridotte assolute". Revista di Grammatica Generativa. (to appear)

Bouchard, D. 1984. On the Content of Empty Categories. Foris Publications.

Bouchard, D. 1985. "A few remarks on past participle agreement in French and Italian", ms., University of Texas.

Bresnan, J. and S. Mchombo. 1985. "Topic, Pronoun and Agreement in Chichewa", ms., Stanford University.

Burzio, L. 1981. "Intransitive verbs and Italian Auxiliaries." Ph.D. dissertation, MIT.

Chomsky, N. 1981. Lectures on Government and Binding. Foris Publications.

Conwell, J. and A. Juilland. 1963. Louisiana French Grammar. The Hague: Mouton & Co.

Fourvières, X. 1941. Grammaire Provençale. Avignon: Maison Aubanel Père.

Kayne, R. 1984. "L'accord du participe passé en français et en italien", ms., Université de Paris VII.

Meyer-Lubke, W. 1955. Grammatica Storica della Lingua Italiana e dei Dialetti Toscani. Torino: Loescher.

Miracle, J. 1980. Gramatica Catalana. Andorra: Editorial Casal I Vall.

Nazari, O. 1900. I Dialetti Italici. Milano: Libraio della Real Casa.

Peruzzi, E. 1959. Problemi di Grammatica Italiana. Torino: Edizioni Radio Italiana.

Remacle, L. 1956. Syntaxe du parler wallon de La Gleize. v. 2. Paris: Les Belles Lettres.

Rohlfs, G. 1969. Grammatica Storica Della Lingua Italiana e Dei Suoi Dialetti. v. 2 & 3. Torino: Giulo Einaudi.

Rossini, G. 1975. Capitoli di Morfologia e Sintassi del Dialetto Cremonese. Firenze: La Nuova Italia.

Stowell, T. 1981. "Origins of Phrase Structure." Ph.D. dissertation, MIT.

Yates, A. 1975. Catalan. London: Hodder and Stoughton.

Thomas D. Cravens

Consonant Strength in the Romance Dialects of the Pyrenees

The Romance dialects on both sides of the French-Spanish border in the Central Pyrenees are characterized - or at least have been characterized until recently - by preservation of the original intervocalic voiceless stops. This situation contrasts sharply with developments elsewhere in Western Romance, where original /p/, /t/, /k/ have at least voiced phonemically (e.g. CAPU > Castilian **cabo**, MATURU > **maduro**, FOCU > **fuego**), and in the most progressive speech types, have often disappeared entirely, especially /t/ and /k/ (MATURU > French **mûr**, SECURU > **sûr**. Actual examples from Bearnese and Upper Aragonese are given in (1), from Elcock (1938), compared with Castilian outcomes (cf. also Rohlfs 1970[2]:130-36).

(1)	Latin	Bearnese	Aragonese	Castilian
/p/	CAPANNA	[kapano]	[kapana]	[kaβaña]
/t/	MATURU	[matü]	[maturo]	[maðuro]
/k/	VERRUCA	[berːüko]	[berːuka]	[berːuγa]

In addition to having preservation of the surds, these dialects have the expected Western Romance development of phonemicized degemination of original /pp/, /tt/, and /kk/. Focusing on the dialect which is best documented, the examples in (2) are taken from Badía Margarit's 1950 study of the speech of Bielsa, on the Spanish side of the border. While Belsetán and Castilian coincide in having simplified original geminate (or long) voiceless stops, they differ in their resolutions of monomoraic /p/, /t/, /k/. Since these latter have not voiced in Belsetán, they have merged with the reflexes of original geminates.[1]

(2)

/pp/	/p/	CAPPELLANU	<	**capellán**	'chaplain'
/p/		NAPU		**napo**	'turnip'
/tt/	/t/	GUTTA	>	**gota**	'drop'
/t/		PRATU	>	**prato**	'pasture'
/kk/	/k/	BUCCA	>	**boca**	'mouth'
/k/		FORMICA	>	**fornica**	'ant'

If these dialects are curious in forming an island of conservatism in the vast area of Western Romance degemination and voicing, the dialect of Bielsa is perhaps most interesting in that, at least as of the late 1940s, it appears that the last remnants of long liquids and nasals were still both underlyingly and superficially long. Unfortunately, Badía does not dwell on this at length, but he does give examples such as **penna, capanna, ninno** (where Castilian has **peña, cabaña, niño**), and **bella, payella, bellota,** with **nn** and **ll** representing **fonemas geminados.** He says that geminate /mm/ is more frequent than /ll/, though he gives only two examples, one of which at least must be discarded from consideration as a diachronic preservation in that it may well be a synchronic surface assimilation: **tammién. Commais** might also be viewed as suspect. If so, from the data available, it might not be safe to assume that an underlying long /mm/ was really possible. Still, even if this must be discarded, we are left with long /nn/ and /ll/ in addition to long /rr/.[2]

(3) /nn/ **penna** 'peña', **capanna** 'cabaña', **ninno** 'niño'
/mm/ **commais** 'comadres', **tammién** 'también'
/ll/ **bella** 'bella', **payella** 'straw basket', **bellota**
(and regular preservation of /rr/: TERRA **tierra,**
etc.)

While the preservation of intervocalic voiceless stops in these dialects has attracted considerable attention in the traditional literature of Romance Linguistics, usually in the form of speculation that Basque substratum is responsible for the failure to voice, Belsetán's maintenance of surface length for some units seems to have gone unnoticed.

The view taken here, especially in light of the fact that none of the substratum hypotheses are truly satisfactory, is that lack of voicing in the stops is best viewed as simple conservatism, of which maintenance of length for liquids and at least /nn/ is another manifestation. From this perspective, the point of interest is to examine possible motivations for the failure of historical degemination to apply to all long consonants. Three questions will be addressed.[3]

(4) a. Why have liquids and nasals not lost length completely?

b. What is the relationship between Belsetán consonant history and that of the rest of Western Romance?

And, for those who find Martinet's chain effects to be of interest,

c. Why were the reflexes of /pp/, /tt/, /kk/ and /p/, /t/, /k/ not kept distinct?

An analysis which incorporates the theoretical concept of relative phonological strength provides a means for answering these and other questions which arise. Though the strength approach seems to be gaining respectability, it is still not widely understood - much less accepted -, so that some elucidation of basic precepts might be helpful in understanding what is to follow.

The strength principles adopted in this study are basically those of Foley (1977), though there are a number of differences stemming from one crucial distinction. Whereas Foley allows no phonetic considerations in the heuristic procedures for determining hierarchical relationships, it is assumed here that strength hierarchies essentially are determined by the phonetic reality of the speech chain. There are many repercussions to assuming a phonetic base in a theory of relative phonological strength, perhaps the most important of which is the consideration that some relationships are necessarily universal. Foley assumes that underlying the phonological system of any language there exists a universal network of abstract strength relationships, not phonemically or phonetically pre-assigned, but manifested independently in each case, and thus that all surface instances of strength relationships are language-specific.[4] The approach followed here assumes that there are clear universal phonetic motivations for **some** strength relationships.

For example, accepting that consonant weakening can be characterized as mutation toward the vocalic qualities of full voicing and aperture, then voiced stops are universally weaker than their voiceless correlates by virtue of being in possession of voicing. Likewise, [h] would be universally weaker than [k], since the first has no full occlusion. Geminate voiceless stops would be universally stronger than simple voiceless stops due to their extended hold of occlusion. Other relationships are clearly language-specific. Relative strength within series, such as /p/, /t/, /k/, varies from one language to another, determined by the character of the **Artikulationsbasis** of each language, as yet all but unexplored. It can be seen from American English t-flapping that /t/ is weaker than /p/ or /k/ language specifically. In the periphery of Tuscany, spirantization of /k/ preferentially to /t/ and /p/ shows that the velar is weakest there (cf. Giannelli and Savoia 1978, 1979-80), as it also appears to be in Finnish, where differential consonant gradations obtain regularly: **kita** 'mouth' - **kidat** 'mouths', **ripa** 'door handle' - **rivat** 'door handles', but **sika** 'pig' - **siat** 'pigs' (all nominative). Kinyarwanda (Bantu) has apparent historical spirantization of /p/ to /h/, and of /b/ to /β/, providing evidence that in that language the bilabials are the weakest of the stops (Kimenyi 1979:2-3). Importantly, this last example illustrates the theoretical premise that barring other disturbances, strength relationships remain constant within a given phonological system; in this case, if /p/ behaves as weaker than /t/ and /k/, it is predictable that /b/ will be weaker than /d/ or /g/ in the same language.

This brief introduction omits a number of principles and further clarifications necessary for a full understanding of the theory, but those relevant here will be introduced in the text. Others can be found in Cravens (1984, and in press). Three basic tenets should be made clear at the outset, however. First, as perhaps implied above but not stated outright, sound change, whether introduction of an allophonic canonical form or phonemicization of change, is expected to operate preferentially, and to generalize sequentially, to classes and within classes, along phonetically-determined parameters. Intervocalic voicing, for example, would be expected to affect first one member of the series /p/, /t/, /k/, then generalize to the others in a principled order. Degemination, which is relevant to the discussion here, is predicted to affect exhaustively one class of units (e.g. voiceless obstruent geminates) before proceeding to another class (e.g. long sonorants). Second, the phonetic manifestations of universal strength hierarchies are at least in part determined language-specifically: one language may have bilabials stronger than either dentals or velars, another may have velars strongest, and so on. The relationships are open to discovery. Third, it is assumed that relative susceptibility to weakening and strengthening is the result of phonetic factors, both articulatory and acoustic.[5]

A full treatment of the questions raised above in (4) requires space not available here, but even a brief glance reveals that the Belsetán outcomes are not peculiar, but rather, that they are simply a conservative subset of the strategies followed **grosso modo** by all of Western Romance, and which are now in the early stages of development in dialects which by any criteria would be classified as Central Italian in type, and thus presumably quite distinct.

It may be helpful to review the data by comparing the diachrony of Belsetán to that of Castilian (5). Latin geminate stops are simplified in both. Single voiceless stops are preserved in Bielsa, voiced (at least) elsewhere. Long /rr/ is maintained in both speech types, but whereas Castilian has palatalized long /ll/ and /nn/ regularly, the evidence indicates that in Belsetán the straightforward internal development is maintained length, now disappearing at least at the surface, and in some cases lexically replaced by Castilian palatalized forms. Again, long /mm/, which has merely simplified in Castilian, may or may not be maintained as a true phonological entity in Belsetán; it may be no more than a manifestation of synchronic assimilation.

(5)

Latin	Belsetán	Castilian
/pp/, /tt/, /kk/	/p/, /t/, /k/	/p/, /t/, /k/
/p/, /t/, /k/	/p/, /t/, /k/	/b/, /d/, /g/
/rr/	/rr/	/rr/
/ll/	/ll/ (~ /ʎ/)	/ʎ/
/nn/	/nn/ (~ /n/)	/ñ/
/mm/	/m/ (assim. [m:]?)	/m/

> "la geminación de fonemas es, entre todos los rasgos fonéticos del belsetán, el que más vacila actualmente; en efecto, es muy frecuente la pronunciación de estos fonemas como simples, de una manera especial **mm** y **l-l**" (Badía Margarit 1950:88).

In attempting to discover how these developments may have come about, it is interesting to compare Belsetán and Castilian with a progressive Italian dialect, the speech of Northeastern Corsica (6), which has at the allophonic stage many of the mutations which Western Romance has phonemicized.

(6)

Latin	Belsetán	NE Corsican	Castilian
/pp/→[pp]	/p/ →[p]	/pp/→[p]	/p/ → [p]
/p/ →[p]	/p/ →[p]	/p/ →[b]	/b/ → [β]
/rr/→[r:]	/rr/→[r:]	/rr/→[r:] ~ [r]	/rr/→[r:]
/ll/→[l:]	/ll/→[l:]~[l]	/ll/→[l:]	/ʎ/ →[ʎ]
/nn/→[n:]	/nn/→[n:]~[n]?	/nn/→[n:]~ [n]	/n̆/ →[ñ]
/mm/→[m:]	/mm/→[m]~[m:]?	/mm/→[m:]~ [m]	/m/ →[m]

In NE Corsican, [p], [b], [w] appear V-V from underlying /pp/, /p/, /b/:

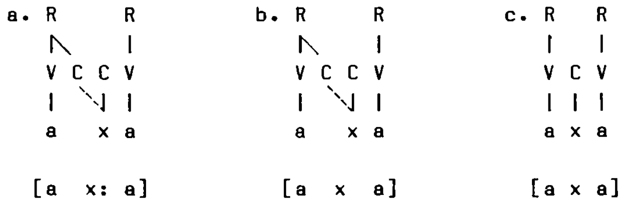
[skapá] 'escape, cf. Italian **scappare**
[ugabu] 'the head', Italian **il capo**
[kawalu] 'horse', Italian **cavallo**

(data from Dalbera-Stefanaggi 1978:56-57)

The surface realizations in Corsican are similar to those of Belsetán, with three exceptions. First, whereas Belsetán has neither diachronic nor synchronic weakening of simple voiceless stops, Corsican has synchronic voicing: anlaut [kabu] but [ugabu] (m. def. art. /u/); [kawalu] but [ugawalu]. Second, Belsetán has merged the reflexes of original /p/, /t/, /k/ and geminate /pp/, /tt/, /kk/ in underlying structure, and therefore also at the surface, whereas Corsican has kept them distinct: the single intervocalic stops are synchronically (allophonically) voiced, and interestingly, underlying bimoraic stops appear as simple units at the surface (/skappa/ ⟶ [skapá] 'escape', /skapa/ ⟶ [skabá] 'behead'). The third major distinction between the two types is found in the historical outcomes of long /ll/ and /rr/. Noting the distinction reveals a curious resemblance, however. NE Corsican and Belsetán both have one long liquid which has maintained distinctive length regularly in the word, and same unit has generalized length to word initial position as well. The difference is that in Belsetán it is /rr/ which fits this description, while in NE Corsican it is /ll/. Corresponding to Belsetán (and Castilian) [r:] in, e.g. [r:ico] 'rich', is NE Corsican [l:], thus anlaut-conditioned [l:ume] 'light', but [ulume] 'the light'. In both languages these are the only members of the liquid and nasal grouping to preserve distinctive length consistently. For the other liquids and nasals in both there is variation in actual speech, perhaps indicating incipient loss of (at least) surface length.

Considering first just the outcome of the original long or geminate stops in Belsetán and NE Corsican, these can be described synchronically in nonlinear terms, as in (7b) and (7c). (7a) shows one possible interpretation (Saltarelli 1983) of Latin geminates (which also applies to modern Italian, very closely cognate to Corsican).

(7)



The syllable structure is bimoraic, the empty node links rightward according to the Italian version of the empty node convention (Ingria 1980), and the surface interpretation is long or geminate. There are some possible problems with this version, but though they are of interest here, they are not crucial to the thrust of the paper.[6] (7b) illustrates the same derivation, with the surface interpretation rule that bimoraic CC is phonetically short. This applies to the underlying long stops of NE Corsican. (7c) represents Castilian and Belsetán voiceless stops; the structure is monomoraic and the derivation is straightforward.

The same derivations apply to the liquids and nasals, but once they are included, it is necessary to list somehow the parameters of application.

(7a) describes:
- /rr/ in Belsetán, Castilian; NE Corsican variably with b.
- /ll/ in NE Corsican; Belsetán variably with b.
- /nn/ in Belsetán and NE Corsican variably with b.
- /mm/ in NE Corsican variably with b.

(7b) describes:
- /rr/ in NE Corsican variably with a.
- /ll/ in Belsetán variably with a.
- /nn/ in Belsetán and NE Corsican variably with a.
- /mm/ in NE Corsican variably with a. (also in Belsetán?)

(7c) describes:
- /m/ in Castilian, Belsetán.

(7a) will handle /rr/ in Belsetán and in Castilian, /ll/ in NE Corsican, and is applied variably with (7b) for /rr/, /nn/, and /mm/ in Corsican, and /ll/, /nn/, and perhaps /mm/ in Belsetán. (7c) applies consistently to the monomoraic /m/ of Castilian, and if Belsetán has lost length for original /mm/, it would apply to its modern reflexes as well.

Presumably, to describe the liquids and nasals in Belsetán, we could establish an analysis such as 7a, with restrictions on the surface interpretation of bimoraic clusters so that /rr/ would be consistently long, /ll/, /nn/, and perhaps /mm/, variably so. But the listing necessary in (7) is at best cumbersome, and it gives no insight as to what systematic principles might be controlling length loss. Moreover, a possibly interesting generalization is being missed, in that we see that NE Corsican behaves so similarly to Belsetán, and that Castilian, while it has not maintained true surface length for /ll/ and /nn/, has nevertheless not just simplified them either. Developments in the three speech types appear to be at least partially congruent, but no motivation for this can be expressed.

In (8) it can be seen that a sort of ordering of historical susceptibility to loss of length in each language is possible. Depending on the status of the reflexes of original /mm/, Belsetán shows either two or three types of outcome: maintenance of length (/rr/), variable surface length (/ll/ and /nn/), and either variable length or loss of length for /mm/. Corsican has two types: maintenance (of /ll/), or variable surface length (/rr/, /nn/, /mm/). Castilian maintains /rr/ regularly, and just as regularly has simplified original /mm/. Again, Castilian outcomes of long /ll/ and /nn/ are interesting in that they have not regularly merged with their simple correlates as the long /mm/ has, but are kept distinct through palatalization.

(8)

Belsetán	NE Corsican	Castilian
/rr/ /ll/ (/mm/)	/ll/ /rr/	/rr/ (/ll/) /mm/
/nn/	/nn/	/nn/
(/mm/)	/mm/	

Focusing closer on just Belsetán and Castilian, it is possible to arrange a provisional scale of susceptibility to length loss for just these two languages, including for both the voiceless stops (9). The long /rr/ is least susceptible to loss of length (and in fact has generalized to initial position in both speech types); according to Badía's report, long /ll/ and /nn/ are next in line in Belsetán. Castilian presents a problem for interpreting the development of long /ll/ and /nn/, but since the two have not simply degeminated and merged with their simple correlates, it might be useful to view these as more resistant to simple length loss than original long /mm/ or the voiceless stops, which have simply degeminated. Belsetán has also simplified /pp/, /tt/, /kk/ phonemically, and if the fate of /mm/ is not totally certain, the fact that it can appear at least as an assimilation suggests that its strength position is midway between the stops and /ll/, /nn/.

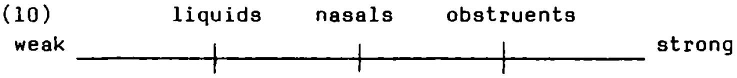
(9) Belsetán and Castilian

/rr/	/ll/ /nn/	/mm/	/pp/ /tt/ /kk/
4	3	2	1

So far we have a description of the differential effect of length loss in Belsetán and NE Corsican, and suspending various reservations, for Castilian as well. But we do not have an understanding of how, or in theoretical terms why, the effects of length loss are arranged as they are in these languages. If we invoke the theoretical concept of relative phonological strength, though, and accept that a reduction in length is a weakening, the scale in (9) can be viewed as the local manifestation, for Belsetán and Castilian, of strength units on the lengthening parameter. That the strength relationships which apparently obtain here are local, that is, language- or language-family specific, is evident when it is considered that NE Corsican has /ll/ as the strength unit least susceptible to length loss, not /rr/ as in Belsetán and Castilian.

This hierarchy is precisely the reverse of that usually proposed in many versions of descriptions of relative strength. In (10) is the hierarchy given in Vogel (1982), which corresponds closely to that of Hooper (1976) and others, this in turn similar to the better known sonority hierarchy, as recently employed by Selkirk (1984). In this scenario, liquids are weakest and obstruents are strongest, whereas, with length loss characterized as a weakening, the analysis conducted here leads to the conclusion that in length loss, the liquids show greatest strength, nasals follow, and obstruents occupy the weakest position.

(10)



Universal strength hierarchy according to Vogel 1982:99

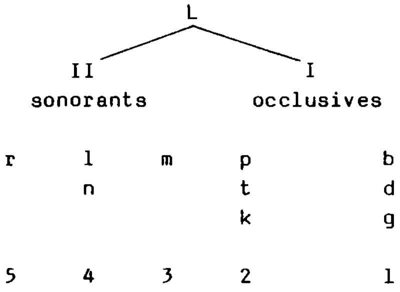
It would appear that the Corsican and Belsetán data contradict Vogel's hierarchy, and that they constitute a contravention of principles which are to some extent established and accepted. But in a more subtle interpretation this need not be the case, for at least two reasons. First, it is to be expected that hierarchies will be process specific, that is, strength relationships will differ according to the phonological process in question, as defined by the relevant environment and the phonological segment or group of segments undergoing the process. Weakenings such as /p/ → [b] or /b/ → [β] are essentially vowel intrusions (substantive reductions of the consonantal articulation; cf. Pagliuca and Mowrey (in press), that is, the extension of the vocalic qualities of voicing in the first instance, aperture in the second, into the acoustic space of the consonantal obstruction, V--->C<---V. It is no accident that voicing and spirantization of stops typically occur with preference intervocalically. Loss of length, or degemination, is essentially (temporal) consonantal reduction, not an extension of vocality, since degemination by definition does not occur in true intervocalic environment; if we accept the structure VCCV, neither C is surrounded by vowels, V--->CC<---V. In syllabic terms, voicing or spirantization has no effect on the skeletal structure of the word, while degemination or length loss, once it occurs at the underlying level, is a deep restructuring, CC ===> C. Banal though these observations are, they are nevertheless essential to understanding that a strength hierarchy which may hold for intervocalic weakening in a language may not hold for a process such as degemination.

Second, phonetic principles impugn the validity of assigning sonorants and stops to the same monolithic strength hierarchy automatically. In the case at hand, the phonetic nature of liquids and nasals as opposed to occlusives might be seen as a concrete motivation for their assignment to distinct phonetically-defined strength parameters, as well as for their discrepancy in phonological behavior with reference to length loss. While a geminate or long stop is basically a hold or delay of offset, e.g. Italian **fat-to**, a geminate or long liquid or nasal is continuous, as in Italian **fanno.**[7] The syllable break in a held occlusive is discrete. For liquids and nasals it is not.

It is expected that strength parameters will be phonetically defined, so that liquids and nasals will typically form one subparameter characterizable as that of the vocalic (or sonorant) consonants, while the stops form another. These may conjoin, as they appear to do in describing Belsetán and Castilian susceptibility to length loss, but this is determined language-specifically and process-specifically.

A more detailed notion of the parameters involved in the Belsetán data is given in (11). Here, sonorants (or vocalic consonants) and occlusives are two subparameters which incidentally conjoin to form the inclusive parameter of Belsetán susceptibility to length loss. The strength relationships are implicational: those strength units with lower values weaken (lose length or degeminate) preferentially to those with higher values. Strength units **p**, **t**, and **k**, for example, are more suspectible to length loss than **m**, which is more vulnerable than **l** or **n**, these in turn weaker than **r**.

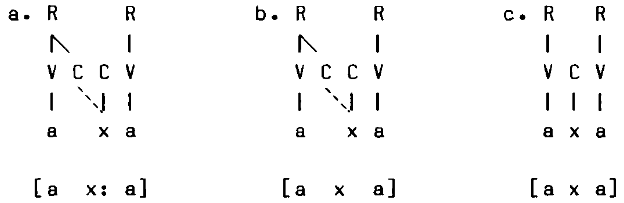
(11)



With this hierarchy established, it is possible to describe Belsetán diachronic length loss in a multiliner derivation, aided this time not by outright listing, but instead by an implicationally determined rule referring to indexed phonological strength.

At the very earliest stages it is safe to assume that Belsetán, or the dialects from which it developed, had the same sort of geminate structure as Latin or modern Italian (12a). If this version is acceptable, the empty node reassociates to the right, and geminates result for occlusives, liquids, and nasals. If the present state of NE Corsican can be taken as an apparent time reflection of the real time development of Belsetán - admittedly a speculation, but a useful one -, the occlusives then would have lost length at the surface, perhaps variably at first, as in (12b), and, without weakening of the original single /p/, /t/, /k/, the two series eventually would have merged, as they did (12c). Today's liquids and nasals in fact, the last to feel the effects of any sort of degemination, have reached the stage described by (12b), where the likelihood of a long surface manifestation is describable with reference to relative position on the lengthening strength parameter (11).

(12)



Belsetán: Bimoraic consonants are long if $S \geq 4 \sim 5$.

The surface interpretation rule states that identical CC clusters at the syllabic level are realized as phonetically long at the surface if their strength values are equal to or greater than 4, varying with 5. Incidentally, the facts for Castilian can be described similarly (with, however, necessity for further specification, by dialect, of actual pronunciations of the fused bimoraic lateral and nasal).

Castilian: Bimoraic consonants are long if $S \geq 5$, fused if 4.

This, then, provides a description of how, and a theoretical means of understanding why, the liquids and nasals have not lost length altogether in Belsetán. Degemination, whether in the form of a surface rule, or syllabic restructuring, has not reached the higher end of the strength scale consistently, and has not reached the strongest member, /rr/, at all; if anything, this strongest unit has generalized its strength further, to initial position. The development is not haphazard, but describable in coherent terms, and the three questions raised at the outside in (4) can now be addressed.

a. **Why have liquids and nasals not lost length consistently when other geminates have simplified?** In the strictest sense, the question cannot be answered. None of the principles adopted here provide a means for determining in any meaningful way why the liquids and nasals have not simplified at the syllabic level, and therefore at the surface as well, as they did in Portuguese, for example. Moreover, a direct answer which invokes their greater phonological strength relative to obstruents as the cause of their resistance to complete length is circular, since their preservation is the evidence of superior phonological strength.

But strength principles do enable an understanding that as consonant length disappears from the phonological system, it is expected that it will be lost sequentially rather than system-wide, and that this gradual loss will progress along parameters which can be identified and motivated phonetically. Strength theory springs from the prime observation that sound change typically applies selectively to a phonological unit or class of units, and then generalizes. Both units and classes of units are definable (or at least characterizable) phonetically, and these phonetic qualities determine relative susceptibility to mutation, depending on the relevant process. In the specific case of surface length loss (eventually resulting in underlying length loss), the continuous vowel-like quality of liquids and nasals assures that a block on a phonetically bisyllabic articulation can be expected to apply last to them, the units with the least clear syllabic break. The period between closure and release which characterizes geminate stops, and identifies the syllable boundary, is absent in long liquids and nasals. In sum, it is not odd, but rather to be expected, that a mutation will operate selectively. It if further expected that selection will be made process specifically, in accordance with phonetic principles.

b. **What is the relationship between Belsetán consonant history and that of the rest of Western Romance?** Posner (1972:48-49) has pointed out that the failure of /p/, /t/, and /k/ to voice is troublesome only if a fully homogeneous earlier stage of Western Romance is assumed. If this demand is eased, as the Pyrenean dialect data suggest it should be, Belsetán developments may remain as perhaps curious, but not anomalous. With regard to the original voiceless intervocalic stops, Belsetán simply has not undergone any type of weakening. There is no mystery here; Belsetán is merely conservative, a fact also reflected in maintenance of initial /f/, and in the continued possibility for surface length deriving from original /ll/ and /nn/.

Loss of length for original /pp/, /tt/, and /kk/ in Belsetán coincides with developments in Castilian. Consistent maintenance of length of /rr/, and its further generalization to word initial position, also parallels the Castilian outcome. The two speech types find a distinction in historical treatment of geminates only in the cases of /ll/, /nn/, and perhaps, /mm/. Yet even here, a pattern is evident: the units which in Belsetán are most subject to realization as long at the surface are precisely those which in Castilian have not simply degeminated, but instead have maintained a strengthened articulation through palatalization. This is not coincidental, but motivated in terms of relative phonological strength.

Belsetán and Castilian have both maintained length consistently for just one strongest strength unit, long /rr/. At a more abstract level, the other members of the strenght hierarchy appear to coincide as well. Castilian's failure to simplify original /ll/ and /nn/ could be viewed on independent grounds as evidence that these are stronger than regularly simplified /mm/, and this interpretation is upheld circumstantially by the fact that the same relationships appear to obtain in Belsetán. The two closely cognate dialects manifest identical strength relationships, then: /rr/ is for both the most resistant to length loss, followed by /ll/ and /nn/ in no identifiable order, and then /mm/. And both have the obstruents as preferentially subject to length loss underlyingly as well as at the surface. In degemination, as well as in the development of original /p/, /t/, /k/, Belsetán simply has been more conservative than Castilian, finding itself only recently in the stage of resolving the loss of length for /ll/ and /nn/.

c. **Why were the reflexes of /pp/, /tt/, /kk/ and /p/, /t/, /k/ not kept distinct in Belsetán?** The implication of the question, that they 'should' have been kept distinct, describes a pseudo-problem. Even a cursory glance at just a few languages shows that though the hypothesis of therapeutic avoidance of merger may be an appealing notion to the linguist, unbroken chains of native speakers pass on their evolving language with no apparent necessity to resort to it. Though individual lexical items may have exceptional developments to avoid unfortunate mergers, such as the well-known failure of diphthongization and syncope in EMERITAM > **Mérida**, it would appear that phonological systems are not subject to therapeutic correction in response to a **Verständigungsprinzip der Sprache** (Weinrich 1958). French and English are notorious for their homonyms, and the list need not be confined just to pairs such as **meet**, **meat**, **peut**, or even **poids**, **pois**, where syntax and context of situation presumably reduce greatly the danger of confusion. Wright (1982:51) cites the case in French in which for those speakers who have no surface contrast of [e] and [ɛ], [ʃã:té] can represent **chanter**, **chantez**, **chanté**, **chantée**, **chantés**, **chantées**, **chantai**, **chantais**, **chantait**, or **chantaient**. Martinet's chains are after-the-fact descriptions of possible, indeed common, developments, but they are not reliable predictors or consistent theoretical motivations of them. The teleological power ascribed to systematic avoidance of merger is a mirage.

For the specific problem here, reflection on the character of strength parameters as posited provides a linguistic motivation for the failure of degemination of /pp/, /tt/, /kk/ to have a push-chain effect on /p/, /t/, /k/. Stated simply, degemination and intervocalic weakening are two different phonological mutations, operating along two distinct strength parameters describing separate implicational domains, these not necessarily linked. Degemination and single-stop weakening can coexist, and they do in various ways, for example, in many varieties of Corsican, one of which we have seen for the Northeastern dialects. An historical degemination can also feed an eventual monomoraic weakening, as it does in varieties of Canary Islands Spanish with allophonic voicing, so that e.g. original BUCCA 'mouth' has develop-

> ed as modern [boga], alongside PAUCU > [pogo] (Oftedal 1985:74). But there is no linguistic reason to assume that the two **must** be implicationally associated. The geminate stops merged with their single correlates in Belsetán simply because the latter underwent no change.

If this analysis is acceptable, it has been shown that the Belsetán peculiarities of preserved intervocalic stops and variably maintained surface length for liquids and nasals (except for constant [rː]) can be described coherently, as well as theoretically motivated, with the incorporation of principles of relative phonological strength, these rooted in a phonetic base. It has also been shown that the close congruence of Belsetán and Castilian outcomes is not haphazard, but that it follows from consideration of the same principles. In this and other cases, application of the concept of relative phonological strength permits theoretical understanding which is not possible otherwise. No claim is made here that a full-fledged strength theory can or should replace other approaches. But by virtue of allowing perception of order where once there was none, it does appear that precepts of implicational strength are a welcome addition to the linguist's set of heuristic tools and explanatory devices.

NOTES

[1] As an ecclesiastical term subject to the vicissitudes of learned vocabulary, **capellán** is quite likely not a direct illustration of popular development. It is the only secure example of /pp/ > /p/ which Badía gives, though, and it is safe to assume that it coincides with genuine Belsetán outcomes of earlier /pp/.

[2] Mott's more recent field research in the neighboring dialect of Gistaín revealed no preservations of length for /ll/ or /nn/, but simplification, e.g. **belota**, **nino**, reflecting a more progressive stage still distinct from Castilian developments. Long /rr/ is generally maintained in **chistavino** (Mott 1984:114-17).

[3] A brief review of the various substratum explanations is available in Jungemann (1955:239-40), to which should be added Politzer (1954). All of these hypotheses are based on the fact that Basque, which toponymic evidence suggests was once the language of the area (Lapesa 1981[9]:29-32;

Menéndez Pidal 1968), has preserved intervocalic surds in early Latin borrowings, e.g. PACE > [pake], along with simplified voiceless geminates as in PECCATU > [pekatu] (Michelena 1961). On the face of it, then, a substratum hypothesis might seem to be a plausible account. But even if we unwisely fail to heed Blaylock's (1966:431) warning that lack of change requires no external causation to motivate it, the claim of Basque influence is fraught with other difficulties, some inherent to any substratum explanation, others specific to this problem. Wanner (1977) has pointed out that any evolutionary change (in this case, lack of change) which a substratum language is able to produce in the conquering language must necessarily be a possible change in the language irrespective of substratum effects. Thus there is no procedural device for determining that a given change **must** be attributable to the effect of the substratum language, and chance coincidence cannot be distinguished from causal relationship. A related obstacle to the coherence of substratum hypotheses is that any substratum account by definition describes non-uniform causation (Naro 1972); all substratum hypotheses assume that only some, not all, of the substratum language's peculiarities are passed on. An example pertinent here is the fate of word-initial /f/. The spirantization which eventually led to loss (e.g. FILIU > [ixo]) has also been claimed to be the result of Basque substratum (cf. Lapesa 1981[9]:38-39), another immediately plausible account in that at an early stage Basque appears not to have had a word-initial /f/, modern Basque forms with initial [f] seem to be recent introductions, and earlier borrowings from Romance have either no remnant of /f/ (e.g. FICU > iko) or substitution with /b/ or /p/ (FAGU > bago, FESTA > pesta) (Michelena 1961:262-65). Yet in the Pyrenean dialects which have preservation of /p/, /t/, /k/ intervocalically, initial /f/ has not been lost: Belsetán /fornika/ < FORMICA, /fondo/ < FUNDU, /feito/ < FACTU, etc. compared to Castilian /ormiga/, /ondo/, /ečo/. Viewed conjointly, the two substratum hypotheses present a quagmire of inconsistency. If the assumptions about early Basque phonology are correct, substratum effect is plausible in both cases. Taking /fornika/ and /ormiga/ as representative, however, the best case that can be made is that substratum effects **may** result in a disturbance, as in loss of initial /f/ in Castilian, and preservation of /k/ in Belsetán; they also **may not** result in a disturbance, however, as in Belsetán preservation of /f/ and Castilian voicing of /k/ > /g/. A substratum effect can act or fail to act unpredictably, then, and once again, a crucial test case cannot be established. Any congruence of the phonology of the substratum language and that of the language which replaces it can be claimed to have been caused by substratum influence as long as the known chronology does not preclude it. No refutation of the hypothesis is possible, leaving not a substratum theory, but a substratum doctrine, to be accepted or rejected as a matter of faith, by appeal to authority, etc.

[4] Misunderstanding of this point has often triggered premature dismissal of Foley's precepts. He makes it clear that in his view, while the phonetic realizations of universal strength elements may incidentally coincide in some languages, they are determined language-specifically. This is expressed in the particular manifestation principle: "Phonological elements may receive different phonetic manifestations in different languages" (Foley 1977:49). To be fair to those accused of having misunderstood, however, it must be admitted that even in a very sympathetic perusal of Foley's book, the reader is hard put not to find numerous lapses in consistent application of this principle. The recognition that some strength relationships are universally motivated on phonetic grounds removes the source of the confusion.

[5] Though synchronic mutations can be described adequately just in terms of articulatory facts, they necessarily come about over time, and thus would seem to be the result of complex interaction of acoustic and articulatory considerations. The fiction of absolute synchrony appears to be untenable in the strength approach, which necessarily concentrates on the dynamic character of phonological alternations.

[6] One of the difficulties encountered in this analysis is that synchronic derivation of e.g. Italian **notte** 'night' is conflated with the historical development NOCTE > **notte**. The empty node interpretation as in (7a) can be seen as a satisfactory account of the history of the assimilation of /k/ to /t/, with the resulting geminate. It also provides a clear motivation for, and description of, Italian **raddoppiamento sintattico**, the phenomenon of synchronic word-initial lengthening as in /tre#kase/ → [trek:ase], as opposed to /la#kasa/ → [lakasa]; words which cue lengthening in the

breath group, such as **tre**, would end with an (etymologically motivated) consonantal empty node which reassociates with the following C. Without independent evidence, however, it is not clear that this analysis should be extended automatically to synchronic derivation of intraword geminates, where there is no surface evidence to argue for preference of an empty node interpretation over one of direct association of CC to the surface.

[7] Selkirk (1984) has recently argued forcefully against the utility of major class features, among others, [+sonorant], in making the case for analysis in terms of a sonority hierarchy. Chomsky and Halle judge the feature [+sonorant] redundant in English (1968:85), and classify English liquids as both [+vocalic] and [+continuant], while [n] and [m] are distinguished only by [+nasal] (1968:177). The ability to separate liquids and nasals from stops, fricatives, etc. is useful here, though, if not for purposes of traditional binary analysis, for cogent description of their common phonetic qualities, reflected in their similar strength behavior.

REFERENCES

Badía Margarit, A. 1950. El habla del Valle de Bielsa (Pirineo Aragonés). Barcelona: Instituto de Estudios Pirenaicos.

Blaylock, C. 1966. Assimilation of stops to preceding resonants in Ibero-Romance. Romance Philology 19.418-34.

Chomsky, N. & M. Halle. 1968. The sound pattern of English. New York: Harper and Row.

Cravens, T.D. 1984. Intervocalic consonant weakening in a phonetic-based strength phonology: Foleyan hierarchies and the **gorgia toscana**. Theoretical Linguistics 11.269-310.

Cravens, T.D. in press. The syllable and phonological strength: gradient loss of gemination in Corsican. To appear in Papers from the 7th International Conference on Historical Linguistics, ed. by Anna Giacalone Ramat, Onofrio Carruba, and Giuliano Bernini. Amsterdam: John Benjamins.

Dalbera-Stefanaggi, M.-J. 1978. Langue corse: une approche linguistique. Paris: Klincksieck.

Elcock. W.D. 1938. De quelques affinités phonétiques entre le béarnais et l'aragonais. Paris: Droz.

Foley, J. 1977. Foundations of theoretical phonology. Cambridge: Cambridge University Press.

Giannelli, L. and L.M. Savoia. 1978. L'indebolimento consonantico in Toscana, I. Rivista italiana di dialettologia 2.23-58.

Giannelli, L. and L.M. Savoia. 1979-80. L'indebolimento consonantico in Toscana, II. Rivista italiana di dialettologia 3-4 (numero unico).38-101.

Hooper, J.B. 1976. An introduction to natural generative phonology. New York: Academic Press.

Ingria, R. 1980. Compensatory lengthening as a metrical phenomenon. Linguistic Inquiry 11.465-97.

Kimenyi, A. 1979. Studies in Kinyarwanda and Bantu phonology. Edmonton: Linguistic Research, Inc.

Jungemann, F.H. 1955. La teoría del sustrato y los dialectos hispano-romances y gascones. Madrid: Gredos.

Lapesa, R. 1981[9]. Historia de la lengua espanola. Madrid: Gredos.

Martinet, A. 1952. Celtic lenition and Western Romance consonants. Language 27.192-217.

Menéndez Pidal, R. 1968. Toponimia prerrománica hispana. Madrid: Gredos.

Michelena, L. 1961. Fonética histórica vasca. San Sebastián: Diputación de Guipúzcoa.

Mott, B.L. 1984. Vocalismo y consonantismo del chistavino (el habla de Gistaín, Pirineo Aragonés). Folia Phonetica (Revista del Laboratori de Fonètica "Pere Barnils") 1.107-31.

Naro, A.J. 1971. On 'f >h' in Castilian and Western Romance. Zeitschrift fur romanische Philologie 88.459-62.

Oftedal, M. 1985. Lenition in Celtic and in Insular Spanish. Oslo: Universitetsforlaget.

Pagliuca, W. and R. Mowrey. in press. Articulatory evolution. To appear in Papers from the 7th International Conference on Historical Linguistics, ed. by Anna Giacalone Ramat, Onofrio Carruba, and Giuliano Bernini. Amsterdam: John Benjamins.

Politzer, R. 1954. On the development of Latin stops in Aragonese. Word 10.60-65.

Posner, R. 1972. Positivism in historical linguistics. Readings in Romance linguistics, ed. by James M. Anderson and Jo Ann Creore, 39-51. The Hague: Mouton.

Rohlfs, G. 1970[2]. Le gascon: études de philologie pyrénéenne. (= Beiheft ZrPh, 85). Halle: Niemeyer.

Saltarelli, M. 1983. The mora unit in Italian phonology. Folia Linguistica 17.7-24.

Selkirk, E. 1984. On the major class features and syllable theory. Language sound structure (Studies in phonology presented to Morris Halle by his teacher and students), ed. by Mark Aronoff and Richard T. Oehrle, 107-136. Cambridge, Massachusetts: The MIT Press.

Vogel, I. 1982. La sillaba come unità fonologica. Bologna: Zanichelli.

Wanner, D. 1977. The heuristics of substratum. Studies in Romance linguistics, ed. by M.P. Hagiwara, 1-13. Rowley, Massachusetts: Newbury House.

Weinrich, H. 1958. Phonologische Studien zur romanischen Sprachgeschichte. Munster: Aschendorff.

Wright, R. 1982. Late Latin and Early Romance. Liverpool: Francis Cairns.

Jorge M. Guitart

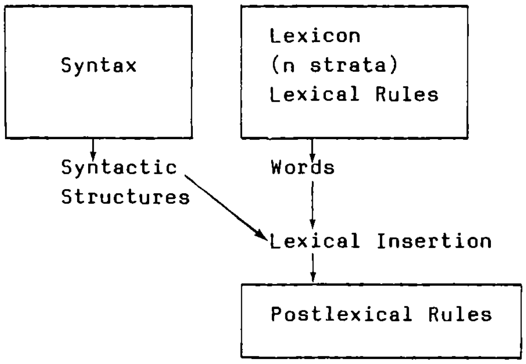
The Case for a Syntax-dependent Postlexical Module in Spanish Phonology

I would like to sketch here a hypothesis concerning the structure of the postlexical component of the phonology. The data are from Spanish in general and from the Cibaeño dialect of Dominican Spanish in particular. The paper revises my 1985 treatment of aspiration and liquid gliding in Cibaeño, which was done within the framework advanced by K.P. Mohanan and Tara Mohanan in their 1984 treatment of nasal assimilation in the Dravidian language, Malayalam. In that seminal paper, the Mohanans incorporated a significant modification into the model known as Lexical Phonology.

In Lexical Phonology - the program of research developed in the last few years by Paul Kiparsky, K.P. Mohanan and others - a distinction is made between a lexical and a postlexical level for phonology. The lexical level is associated with the lexicon. Together they constitute the lexical module. At the lexical level, all phonological phenomena are word-bounded and many occur as a consequence of the contact between morphemes when words are formed, and of the contact between words when compounds are formed. Word stress applies after word formation and so do other rules whose domain is the word. After compound formation, there applies compound stress and all other rules whose domain is the compound word. It is at the lexical level that all rules thought of as morphophonemic in traditional phonology apply. Or putting it differently, all such rules are contained in the lexical module. It should be added that the lexical module contains different strata and that certain processes that apply in a given stratum may not apply in another stratum. The details do not concern us here.

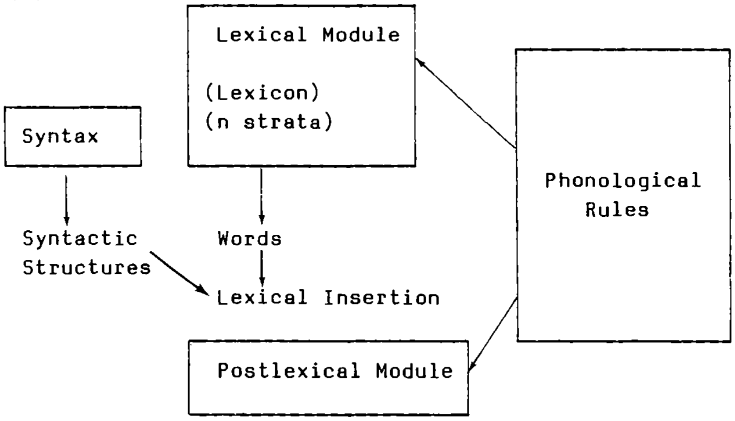
The output of the lexical module is words, which are inserted in the syntactic strings that are the output of the syntax. These strings in turn are the input to the postlexical module where all and only postlexical rules apply. These are the rules thought of traditionally as purely phonological. These rules have no access to either lexical or syntactic information. In the early model of lexical phonology (see, e.g., Kiparsky 1982), all lexical rules are ordered before all postlexical rules. No postlexical rule applies in the lexical module and no lexical rule applies in the postlexical module. A sketch of this model is shown in (1).

(1) Early model of Lexical Phonology:


Syntax
Lexicon
(n strata)
Lexical Rules
Syntactic
Structures
Words
Lexical Insertion
Postlexical Rules

The Mohanans showed conclusively that in Malayalam nasal assimilation must apply both in the lexical and postlexical modules. This led them to propose that phonological rules are not properties of modules but of the grammar. That is to say, there are no lexical rules or postlexical rules; instead, some rules apply only at the lexical level, others at the postlexical level and still others may apply at both levels. A sketch of the Mohanan's model is shown in (2):

(2)


Lexical Module
(Lexicon)
(n strata)
Syntax
Phonological
Rules
Syntactic
Structures
Words
Lexical Insertion
Postlexical Module

In support of the Mohanan's proposal I argued in Guitart 1985 that /s/ aspiration (general to Caribbean Spanish, including Cibaeño) must apply at both the lexical and postlexical levels, and so must Cibaeño liquid gliding, except that the latter rule does not apply to clitics at the lexical level. (All Cibaeño data are from Alba (1979, 1984).)

The relevant facts are as follows. Postnuclear /s/ may be aspirated in both clitics and words, e.g., loh.kóh.toh for **los costos** 'the costs', but prenuclear /s/ is not aspirated, e.g., é.sa for **esa** 'this one' (*é.ha). However, in normal speech, the glottal fricative is in the onset postlexically if a vowel follows, e.g., lo.há.bre for **los abre** 'he opens them', dó.ho.tréh.bán for **dos o tres van** 'two or three are going'. Therefore, in the latter two cases aspiration must have applied to **los** and **dos** at a level where /s/ was postnuclear. I assumed that, in the same cases, resyllabication applied only after bracket erasure and that aspiration applied before bracket erasure, with word- or clitic-final /s/ being indeed postnuclear inside a bracket. Therefore aspiration applied at the lexical level. But given that aspiration is a variable rule and is a process seemingly influenced by the position of /s/ within an utterance (as, for instance, Terrell (1979) has shown), I assumed that aspiration also applied at the postlexical level, where the environment is formulated exclusively in phonetic terms, including the phonetic makeup of neighboring sounds.

As to liquid gliding, word-final /l,r/ in stressed syllables may be realized as the palatal glide regardless of what follows: éy dá for **el da** 'gives', éy ábla for **el habla** 'he speaks', séy dúro for **ser duro** 'to be hard', séy álto for **ser alto** 'to be tall'. In contrast, clitic-final /l,r/ - always in unstressed syllables - become glides only before consonants, never before vowels, and in the latter case are in the onset. Compare ey.dá.to (for **el dato** 'the datum') and poy.pó.ko (for **por poco** 'almost') with e.la.mí.go (for **el amigo** 'the friend') and po.ré.so (for **por eso** 'because of that'). Despite the reference to stress, liquid gliding is actually independent of stress, since it also applies to liquids that are in unstressed syllables word-internally, e.g., kay.bón for **carbon** 'coal', pay.mí.ta for **palmita** 'little palm'. The reference to stress in word-final position is necessary, for if the syllable is unstressed the liquid is deleted rather than realized as a glide, e.g., kánse for **cáncer** 'cancer', fási for **fácil** 'easy'. (In

Guitart 1980, I argued that liquid gliding does apply at some nonsurface level to words like **cáncer** and **fácil**, and then the glide is deleted so that the word conforms to Spanish morphological structure which disallows both glide-ending paroxytones (*kán.sey) and tautosyllabic sequences of high vowel and high glide (*fá.siy). Where glide deletion would apply in the model sketched in this paper is an interesting question, but I will not explore it here).)

Since the rule of liquid gliding obviously distinguishes between clitics and words, I assumed that this must take place at a level where items are bracketed, with the rule applying to word-final liquids because they are postnuclear at that level, and with resyllabication applying after bracket erasure. On the other hand, since preconsonantal clitic-final liquids can be affected, this must take place at a level where no distinction can be made between clitics and words and where the structural description of the rule is formulated exclusively on phonetic terms, and that is the postlexical level. Therefore, liquid gliding applied both at the lexical and postlexical levels.

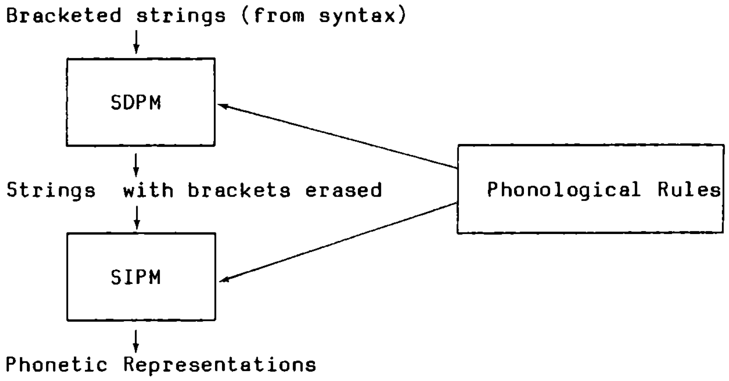
Unfortunately the proposal just sketched has a serious flaw and is in need of revision. The revision involves proposing a new structure for the postlexical level, retaining nonetheless the Mohanans' assumptions that rules are independent of modules and that rule application may be multi-modular.

First of all, the word at the lexical level cannot be the domain of application of either aspiration or liquid gliding. Otherwise, for instance, nouns that in the singular end in /s/ or in a liquid, would show in the plural the aspirated or glided versions of those segments, since in plural formation the plural morpheme is attached to the singular, which is a word. Cibaeno speakers would sometimes pronounce, e.g., *bóheh for **voces** 'voices'- from singular /bos/, **voz** - and *papéyes for **papeles** 'papers' - from singular /papel/; but the plurals are never pronounced that way. Therefore /s/ aspiration and liquid gliding actually do not apply at the lexical level. (I owe this observation to James Harris.)

However, there must be another level in the phonology where rules using syntactic information may apply cyclically to bracketed items. Consider the following facts. All Spanish dialects, including Cibaeno Spanish, have a rule that deletes /s/ from the first person plural suffix **mos** of verb forms if and only if a) the mood is imperative (i.e., the form is a command) and b) it is part of a conjugated-clitic construction (née reflexive or pseudoreflexive). For instance, compare **mirémonos** 'let us look at ourselves' (from **[miremos]+[nos]**) with **mirémoslo** 'let us look at him' (from **[miremos]+[lo]**); and **casémonos** 'let's get married' (from **[kasemos]+[nos]**) with **casémoslos** 'let's get them married' (from **[kasemos]+[los]**). In dialects with second person plural informal forms (i.e., **vosotros** forms) the process is more general as it deletes the final coronal of commands that come into contact with conjugated clitics; compare for example **casaos** '(you, pl., inf.) get married' (from **[kasad]+[os]**) with casadlos '(you, pl., inf.) get them married' (from **[kasad]+[los]**). (There is at least one verb that must be marked lexically as not undergoing deletion in second person plural commands: for **irse** 'to leave', the command is **idos** '(you, pl., inf.) leave' instead of *íos.).

Now, since mood and agreement are not lexical, coronal deletion in command forms must apply outside the lexical phonology but within a module where constituents are bracketed. This I will call the syntax-dependent post-lexical module (SDPM). Its input is bracketed syntactic strings that are the result of lexical insertion, and its output (with brackets having been erased) is the input to a syntax-independent post-lexical module (SIPM), which is the same as the postlexical module of standard Lexical Phonology as revised by the Mohanans. This is sketched in (3):

(3)



I would now like to propose that, similarly to the lexical module, the SDPM has strata. It has at least two strata. In Stratum 1, there occurs the process of clitic-word compounding, which has phonological consequences. That is to say, there are phonological rules that apply as the result of the syntactic contact between a word and its clitics. One of those rules is coronal deletion in Spanish commands. After all applicable rules have applied in that stratum, brackets internal to a word-clitic string are erased, and the word and its clitics constitute a **citicized word,** which is treated as a single word in Stratum 2. This is illustrated in (4) with the derivation of **miraos** 'look at yourselves':

(4) **Stratum 1**

[[mirad] [os]] (From Syntax)
[[mira] [os]] Coronal Deletion
[miraos] (Brackets erased)
↓
Stratum 2

However, coronal deletion does not apply in Stratum 2. For instance it does not apply to a command form that is adjacent to the clitic of another form. Compare (5a) and (5b):

(5) a. **Miraos, digo** 'Look at yourselves, I say'
b. **Mirad, os digo** 'Look, I tell you (pl. inf.)'

At the end of Stratum 1, (5a) would be bracketed as (6a), and (5b) as (6b):

(6) a. [[miraos] [digo]]
b. [[mirad] [osdigo]]

In Stratum 2, rules apply cyclically to ever increasing domains. At the end of Stratum 2, all brackets have been erased and strings enter SIPM.

With respect to the Cibaeño Spanish processes under discussion, I would like to propose that aspiration applies in Stratum 1, but liquid gliding does not. This is intended to characterize the fact that aspiration does not distinguish between clitics and words, but liquid gliding does. Both rules apply in Stratum 2.

Regarding the relationship between syllabic structure and the processes in question, I would like to list in (7) two crucial assumptions of my proposal:

(7) i) syllabication precedes the application of any rules in Stratum 1;
  ii) resyllabication applies automatically at the end of each stratum to any material within brackets.

Consider now the phrases **el oso** 'the bear', **los osos** 'the bears', **dos osos** 'two bears', and **ser oso** 'to be a bear'. In Stratum 1, aspiration applies as the result of clitic-word compounding only to [[los][osos]], resulting in [lo.ho.soh] at the end of Stratum 1, but it does not apply to [[dos][osos]]. Since liquid gliding does not apply in Stratum 1, **el oso** is [e.lo.so] at the end of that stratum. In Stratum 2, aspiration applies again, and this time it can apply to [[dos][osos]], resulting in [dohosoh] after bracket erasure, which is then resyllabified as do.ho.soh.

Liquid gliding applies in Stratum 2 to [[ser][oso]], resulting in [seyoso] at the end of that stratum. (According to Alba (1979) speakers do not resyllabify the glide, but this is of no consequence to our proposal.) Gliding also applies for instance to [el.ko.ko] (**el coco** 'the coconut') resulting in [ey.ko.ko] but it does not apply to [e.lo.so] because the liquid is not postnuclear.

Im summary, I have incorporated into the phonology a syntax-dependent postlexical module, which like the lexical module has strata and in which rules apply cyclically; it is ordered between the lexical module and a syntax-independent postlexical module. Clitic-word compounding takes place in Stratum 1 of the syntax-dependent module. Certain rules, e.g., coronal deletion, apply only in Stratum 1; certain other rules, e.g., liquid gliding, apply only in Stratum 2, but there are rules like aspiration that can apply in both strata. As a final note I would like to add that in the matter of rule applicability to a given stratum or module, I have eliminated the need to invoke the fact that aspiration and gliding are variable. This is as it should be, since the conditions that determine whether a variable process takes place or fails to take place are extragrammatical.

REFERENCES

Alba, O. 1979. Analisis fonológico de /r/ y /l/ implosivas en un dialecto rural dominicano. Boletín de la Academia Puertorriqueña de la Lengua Española 7:1-18.

———. 1984. Estudio sociolingüístico de la variación de las líquidas finales en el español cibaeño. Paper delivered at the 7th Symposium on Caribbean Spanish Dialectology, Florida Atlantic University, Boca Raton, 27-9 April 1984.

Guitart, J.M. 1980. Some theoretical implications of liquid gliding in Cibaeño Dominican Spanish. Paper delivered at the Xth Linguistic Symposium on Romance Languages, University of Washington, Seattle, Washington. (Published in 1981 in H. Contreras and J.Klausenburger, eds., Proceedings of the Tenth Anniversary Symposium on Romance Linguistics, supplement II to Papers in Linguistics 3, University of Washington, Seattle, Washington, 223-28.)

———. 1985. Variable rules in Caribbean Spanish and the organization of phonology. Current issues in Hispanic phonology and morphology, ed. by Frank M. Nuessel, Jr. Bloomington: Indiana University Linguistics Club, 28-32.

Kiparsky, P. 1982. Word formation and the lexicon. Paper delivered at the Mid-America Linguistics Conference, University of Kansas. Published in 1982 Mid-America Linguistics Conference Papers, ed. by Frances Ingemann, Lawrence: Department of Linguistics, The University of Kansas, 1983, 3-29.

Mohanan, K.P. and T. Mohanan. 1984. Lexical phonology of the consonant system in Malayalam. Linguistic Inquiry 15:575-602.

Terrell, T.D. 1979. Final /s/ in Cuban Spanish. Hispania 62:599-612.

Paul Hirschbühler

The Middle and the Pseudo-Middle in French

0.1 Topics discussed

In recent years, there has been considerable interest in the properties of the middle construction in a variety of languages, especially in the Romance languages and in English. By and large, there tends to be a general agreement as to the main properties of the middle in these languages, as well as to the ways in which these properties should be encoded in a grammar. In this paper I will try to contribute to our understanding of the middle in French in two areas.

1. First of all, I will consider some unnoticed facts about the French middle that will allow me to cast some doubts as to the validity of certain aspects of the analysis of the French middle construction as it is generally conceived at the moment. The facts discussed bear on the reality of the syntactic presence of the so-called implicit argument in the middle construction.

2. The second area that I will consider concerns a use of the middle construction that for ease of exposition I have called the **pseudo-middle** construction; it is exemplified in (1):

(1) Cette voiture se conduit d'elle-même.
This car Refl(exive) drives on itself.

The question here is to what extent the similarities between this type of example and the English middle should lead to a parallel analysis, distinct from that of the more typical French middle, whose properties are quite different from that of the English middle.

0.2 Background

Let us first summarize the relevant points of relative agreement.

1. The middle construction is a construction into which a number of **transitive verbs** can enter. The particular verbs allowed vary depending on the language under study. For example, at first sight at least, going from the more restrictive to the less restrictive case, we have the following hierarchy:

English > French > Spanish

2. On the surface, the middle construction is characterized by the fact that an NP which is typically the object of a transitive verb appears as the subject of such a verb, while the NP that would express the theta-role of the subject in the active transitive construction is left unexpressed. In the Romance languages, the verb is in the reflexive form. One very common type of analysis for this construction is along the following lines. a) The reflexive pronoun dethematizes the subject position by absorbing the theta-role that the verb assigns to the external argument. Under this view, the reflexive pronoun is actually considered as the external argument. b) In addition, the reflexive pronoun absorbs the objective case, this resulting in the object moving in the dethematized subject position where it can be assigned Case.

3. While the so-called implicit argument is in some way really present in the Romance languages, this argument is altogether absent from the theta-grid of the middle verbs in English. This is clear from the work of Lakoff (1977) and Van Oosten (1977), and has since been generally adopted, for example in the recent papers by Hale and Keyser (1985) and Jaeggli (1984).

With this as background, I will now turn to the first topic of discussion, that of the presence of the external argument in the syntactic representation of the middle in a language like French.

1. The external argument

The typical facts used to establish the presence of an underlying subject argument can already be found in N. Ruwet's study "Les constructions pronominales neutres et moyennes" (1972). Ruwet used these facts to argue that the middle in French had an underlying agent and that the surface subject was derived by syntactic movement from object position. The possible point of contention is whether these facts really support a view whereby the external argument is projected in the syntax or whether it is, for example, simply present in the theta-grid of the verb, where it might be assigned an **arb** index that would, among other things, prevent it from being projected into the syntax. A number of facts have been put forth to support the view that the external argument is projected into the syntax; I will restrict myself to those that support this position best, i.e., I will examine the case of the so-called agent-oriented adverbs, of the adjuncts predicated of the implicit argument, and of purpose claused controlled by the implicit argument. The argument will have the same structure in each case. I will argue that the examination of a larger array of facts casts some doubts as to the validity of the claim that the types of phrases discussed require the presence in the syntax of the implicit argument.

Secondly, assuming for the sake of discussion that **se** is the external argument, I will point out the fact that **se** cannot be put in relation with any sort of antecedent, contrary what is the case for the other pronouns or empty categories discussed in the literature, raising thereby some more doubts as to the presence in the syntax of the implicit argument.

Let us start with "agent" oriented adverbs.

As is well known, the middle in French allows "agent" oriented adverbs:

(2) Ce livre se lit avec plaisir.
This book Refl reads with pleasure.

The adjunct **"avec plaisir"** describes a psychological state of the reader, and in that sence, one could argue that this phrase would not be allowed in the sentence if at some level of representation the 'reader'-argument would not be present. As examples like (2) are not allowed in the English middle, as can be seen in (3),

(3) *This book reads with pleasure

it is normal to claim that in English the argument that correspond to the typical subject of the transitive use of the relevant verbs is absent. This position has been strongly defended by Lakoff (1977) and Van Oosten (1977), and I think it is correct for English. According to Van Oosten, in English, the surface subject of the middle (the patient-subject, as she calls it) has semantic properties that are prototypical of agent subjects: the entity denoted by the subject, or a property of it, is understood to be responsible for the occurrence of the action denoted by the predicate. Hale and Keyser (1985) and Jaeggli (1984) also hold the position that in the English middle the theta-grid of the verb lacks the external argument of its corresponding transitive use. This allows for the possibility that the implicit argument that is understood to be present in the French middle is either represented in the theta-grid of the verb and is assigned a constant interpretation at that level, or that it is, in addition to being represented in the theta-grid, projected in the syntax. In order to be able to claim that se in the middle is the external argument, one must then show that syntactic constituents that depend somehow on the implicit argument require that this argument be expressed in the syntax proper. Arguments to that effect have been presented on the basis of cases like (2) for example, but the examination of a wider range of facts casts some doubts as to the validity of the conclusion drawn from cases like (2). Consider for example a sentence like (4):

(4) La discussion s'est déroulée avec plaisir/bonne foi/politesse
The discussion went with pleasure/in good faith/courtesy

The reflexive use of the verb here is not the middle one, but, depending on the terminology, the neuter one (Ruwet), the ergative (GB), or the unaccusative use (Relational Grammar)[2]: the verb has only one semantic argument. An example like (4) suggests that "agent"-oriented adverbial phrases like **avec plaisir/avec bonne foi/avec politesse** may be understood as making a predication about an entity not represented syntactically in the sentence, so long as it is present in the theta-grid of the verb as used in that sentence, as it might be the case in the French middle, or so long as the sentence somehow implies that there is an "agent"-participant in the event, as in the case of (4).

The same type of situation can be found with adjectival adjuncts that are understood as predictated of the implicit argument and that are presumably daughters of S, as in (5):

(5) Ce genre de pièce se joue **nu**.
This type of play Refl plays naked.

It is generally considered that an adjunct like **nu** here must be c-commanded by the phrase of which it is understood as predicated. Williams (Class lectures, spring 1985, UQAM) has however argued that in English examples of the passive of the type of (6), the adjective is actually predicated of the surface subject, from which we deduce that the players are nude:

(6) This game is played nude.

There are very clear examples in French where Williams' approach is correct, i.e. the adjunct does not depend on the presence in the syntactic structure of a phrase representing the entity that the adjunct is understood as making a claim about:

(7) a. La prière aura lieu **nu**.
The prayer will take place naked.

b. L'attaque réussira **la tête découverte**: laissez vos casques au camp.
The attack will succeed with the head bare: leave your helmets at the camp.

c. La traversée de l'arctique n'est possible qu'**habillé**.
Crossing the Arctic is only possible dressed.

A third type of fact that has been used to support the view that the external argument of the transitive use is present in the middle comes from control. Examples like (8a), adapted from Demonte (1985) are adduced to support the view that the implicit agent argument of the passive is present in the syntax, as PRO, the subject of the infinitival clause, is understood as coreferential with it: PRO would be given its reference by an argument of the main clause. And one could try to extend this to cases of the middle like (8b):

(8) a. Les coupables furent châtiés pour démontrer le pouvoir de la police.
The guilty people were punished to demonstrate the power of the police.

b. Ce genre de personne s'arrête de temps en temps pour démontrer le pouvoir de la police.
This type of person Refl arrests from time to time to demonstrate the power of the police.

But in French examples at least there is no such coreferentiality implied. Actually, it seems that what is understood as the subject of **démontrer** is the event depicted in the main clause: the arrest-event demonstrates the power of the police. The clearest cases that show that PRO is really controlled by an argument of a higher clause are those where the assumed controller contains a quantifier, as in (9) for example, and where PRO is understood to be under the scope of the quantifier.

(9) Chaque employé est parti pour éviter d'être démis.
Each employee left in order to avoid to be dismissed.

But no relevant case can be constructed either in the truncated passive or in the middle, which does not allow the external argument to be expressed as an adjunct. And in non-truncated passives, when the underlying external argument is expressed in a prepositional phrase, if it is a quantified NP, it cannot be understood as controlling, or having scope over, a PRO in a purpose clause:

(10) a. Chaque candidat a relu le manuscrit pour y repérer le plus de fautes de frappe possibles.

Each candidate reread the manuscript to discover the largest possible number of typos in it.

b. Le manuscrit a été relu par **chaque candidat** pour y repérer le plus de fautes de frappe possibles. The manuscript was reread by each candidate to discover the largest possible number of mistakes in it.

While the (a) example can be understood as meaning that it is each candidate's intention to try to discover the largest possible number of mistakes, this interpretation is impossible in the (b) example. Thus, control facts in French don't support the view that the implicit argument of the middle can syntactically control PRO. I would suggest that in such cases, there is no control: PRO is assigned any sort of constant value, as long as one is able to pragmatically fit it with the semantics of the main clause.

The three cases I have briefly examined are cases where the implicit argument was taken to be the antecent of some phrase. One might wonder whether the implicit argument could itself have an antecedent, as has been suggested is the case for some other implicit arguments. For example, Rizzi (1985) argues that in the Italian examples parallel to the French one in (11), there is a **pro** in object position,

(11) Ceci conduit pro à [PRO conclure ce qui suit]
This leads to conclude the following

and that this **pro** can take an antecedent, as in (12), which is argued to be ambiguous with respect to the reference of **pro**, as indicated in (13):

(12) C'est une illusion de [PRO espérer [qu'un bon repas puisse réconcilier **pro** avec soi-même]].

(13) a. It is an illusion for **x** to hope that a good meal can reconcile **y** with **y**.

b. It is an illusion for **x** to hope that a good meal can reconcile **x** with **x**.

I don't think that anything comparable can be found in the case of the middle. Consider examples like (14):

(14) a. Il est difficile de croire [que ce livre se vende à Pierre].
It is difficult to believe that this book Refl sell to Peter.

b. Je crois que [ce livre s'est vendu à Pierre].
I believe that this book Refl sold to Peter.

c. J'ai de la difficulté à croire que cette voiture se soit enfin vendue.
I have difficulty to believe that this car Refl finally sold.

In the first two cases, the implicit argument of the embedded clause is clearly different from any of the arguments (implicit or explicit) of the main clause, while for the third one, it is compatible with the possibility that the subject of the main clause and the seller are the same person. But there is not indication that one should think that there is coreference, for example expressed by some coindexing; if one adds any sort of quantifier to the imagined controller in the main clause, the interpretation where the implicit argument would be bound by the quantified phrase is excluded:

(15) a. Chaque geôlier espérait que les prisonniers pourraient se fouetter encore longtemps.
Each jailer hoped that the prisoners could Refl whip for still a long time.

b. (Chaque pêcheur+aucun pêcheur n')espérait que les poissons pourraient s'attraper en grand nombre.
(Each fisherman+no fisherman) hoped that fish could Refl catch in large number.

In these cases too therefore we are led to conclude that the implicit argument is not represented in the syntax.

2. The pseudo-middle and verb classes

The middle use of transitive verbs is productive, but not every transitive verb readily accepts the middle construction. The restrictions are however general, even if not perfectly understood. They vary to an extent between the different languages studied. Here I will briefly discuss one particular use of the middle that allows a subset of the verbs that can be used in the regular middle.

Consider the following contrast between the English and the French middle:

(16) a. The book is selling like hotcakes.
b. *The book is buying like it was going out of style.
c. *This knows very easily.

(17) a. Ces livres se vendent partout.
These books Refl sell everywhere.

b. Ce genre de livres s'achète surtout dans les gares.
This type of books Refl buys especially in train stations.

c. Cela se sait depuis longtemps.
This Refl knows for a long time.

d. *Cela se connaît depuis longtemps.
This Refl knows for a long time.

Van Oosten, from whom the examples in (16) are taken, proposes that in English the surface subject of the middle is the deep subject as well, and that as such it must display prototypical semantic properties of subjects: here it will be some sort of responsibility for whatever the predicate describes. With **sell** it is said that properties of the object being sold are construed as responsible for the transaction; the point of not mentioning the seller (or better, eliminating it altogether) is to emphasize the role of the patient as responsible. With a verb like **buy**, it seems that properties of the buyer are considered as having prime responsibility: as such the external argument of **buy** may not be dethematized. Van Oosten argues very convincingly for this point of view. Among other contrasts, she offers the following one as supportive evidence:

(18) a. How dit Alex manage to buy the car? *It's a great car, a real bargain.

b. How did Marie manage to **sell** the car? It's a great car, a real bargain.

The difference between the two examples can be attributed to the hypothesis that with **sell** properties of the patient are taken as relevant, while with **buy** they are not.

Neither French nor the other Romance languages diplay these particular types of restrictions in the standard middle construction. This could be attributed to the fact that somehow the implicit argument is still linguistically present in those languages, with the result that the surface subject is not automatically endowed with prototypical deep subject semantic properties. However, if the facts of English are really related 1) to the elimination of the underlying subject and 2) to general semantic properties of verbs, one might expect to be able to replicate them in languages like French or Spanish if there were ways to further minimize the role of the underlying subject. This is possible to a large extent with the pseudo-middle, which is exemplified again in (19).

(19) a. Cette maison s'est vendue **d'elle-même**[3].
This house Refl sold by itself.

b. Cette voiture se conduit **d'elle-même**.
This car Refl drives by itself.

The typical use of **de soi-même** as far as we know is exemplified in examples like the following:

(20) a. Pierre a fait cela de lui-même.
Peter did this of himself.

b. Pierre est parti de lui-même.
Peter left of himself.

c. La roche est tombée d'elle-même.
The rock fell of itself.

By and large, it seems that **de soi-même** is used to indicate that the responsibility of the event rests with the surface subject of the clause; it denies outside intervention, outside responsibility. We are in a position now to examine

(19a-b). Example (19a) indicates that, from the point of view of the speaker, no outside intervention was necessary for the transaction to go through; (19b) amounts to saying that the intervention of the driver is also null for the driving of the car. Pragmatically, we conclude that the role of the seller or of the driver was as minimal as one can think of in those two cases. Notice that **de soi-même** cannot be used in the passive:

(21) a. *Cette voiture peut-être conduite d'elle-même pendant les premiers deux mille kilomètres. Aprés ça, elle commence à faire des siennes.
This car may be driven of itself during the first two thousand kilometers. After that, it starts to act funny.

b. *Cette maison a été vendue d'elle-même.
This house was sold of itself.

On the basis of examples like (20) and (21) one could claim that **de soi-même** is possible only with deep subjects. However, enough people would point out that verbs like **tomber** are unaccusative verbs, so that this would have to be modified. Let's say that **de soi-même** is possible when the surface subject is the entity to which total responsibility is assigned for whatever event the proposition describes. If this is correct, then one would be led to say that in examples of the type of (19), the **se** does not represent an extra argument, or at least not an argument corresponding to the external theta-role of the relevant verbs.[4] Is there independent evidence for this? It seems there is. Remember the English middle: while verbs like **sell** are perfect, verbs like **buy** are not, because, so the argument goes, with verbs like **buy** responsibility never escapes the buyer, as opposed to **sell** where responsibility may escape the seller. It seems that we find the same with the pseudo-middle in French:

(22) a. Cette maison s'est vendue d'elle-même.
This house Refl sold of itself.

b. ??Cette maison s'est achetée d'elle-même.
This house Refl bought of itself.

(23) a. ?*Cette voiture se lave d'elle-même.
This car washes of itself.

b. *Cette pièce se balaye d'elle-même.
This room Refl sweeps of itself.

c. ?*Ces fruits se mangent d'eux-mêmes (OK: se digèrent...).
These fruits Refl eat of themselves.

d. *Ce tire-bouchon s'utilise de lui-même.
This cork-screw Refl uses of itself.

And in addition, the typical types of tests used to demonstrate the presence of an extra argument in the middle in French are all negative. Let us just look at the "agent" oriented adverbs:

(24) *Cette voiture se conduit d'elle-même avec plaisir.
This car Refl drives of itself with pleasure.

Let us assume then that this construction is parallel to the English middle. One would then say that in these examples there is no implicit argument, and that the surface subject is the deep subject (as in Van Oosten and Hale and Keyser, 1985) or the deep object (as in Jaeggli, 1984).

I would however feel uncomfortable with an analysis that assigns different structures, as opposed to different interpretations, to the middle and the pseudo-middle in French. That is, I would favor an analysis that doesn't eliminate the external argument from the theta-grid, since the external argument must certainly be present in the theta-grid of the standard middle. I think that one could maintain an identical analysis for the middle and the pseudo-middle along the following lines. In the middle, the VP may assign a constructional theta-role to the surface subject, independent of the inherent theta-roles that verbs may assign; this is the constructional AGENT theta-role, which Hale and Keyser (1985) understand as "the entity **responsible** for that which is denoted by the VP". **De soi-même** is associated with the surface subject. Its semantic contribution results in the constructional AGENT role to be emphasized and in the importance of the inherent AGENT role to be downgraded, when the two are not assigned to the same element of the theta-grid. One could then say that in such a case, **avec plaisir** and **de soi-même** are pragmatically incompatible: **avec plaisir** makes the inherent agent role relevant, as we bother introducing characteristics of the agent, while the presence of **de soi-même** results in making the contribution of the agent irrelevant.

Are there independent reasons to consider the pseudo-middle as different from the English middle? The answer is clearly 'yes'.

First of all, although there is a contrast between (22a) and (22b), there are situations where I think (22b) would sound fine. Imagine for example someone with a lot of money, who buys and sells properties very regularly. He may have signed the cheque for the house without really noticing and become aware of the deal only after the money involved had been spent and without the buyer's noticing the expense. At that point, it would be appropriate for the buyer to utter something like (22b).

Secondly, and more generally, it seems that many verbs are allowed in the pseudo-middle that would be excluded in the English middle, for example many of the verbs that refer to mental activity are fine in the pseudo-middle, but excluded in the English middle:

(25) a. Les langues romanes s'acquièrent d'elles-mêmes.
b. Ces théorèmes s'apprennent d'eux-mêmes.
c. Ces formules se mémorisent d'elles-mêmes.

(26) a. *Romance languages acquire easily.
b. *These theorems learn easily.
c. *Those formulas memorize easily.

If those verbs are "agent" oriented, in the sense that **buy** was said to be, then Van Oosten's characterization of the central feature of the English middle would be enough to exclude examples like (26), and the grammaticality of (25) would clearly show the non-equivalence between the English middle and the French pseudo-middle. The pseudo-middle is simply a middle construction where the addition of a phrase like **de soi-même** results in the role of the surface subject as something whose properties facilitate the coming about of the event described by the verb to be highlighted.

4. Conclusion

Let's briefly summarize the main points that have been made.

First, there is no doubt that the middle in French is quite different from the English construction with the same name. In French, contrary to what is the case in English, it is clear that the external argument of a transitive verb is part of the theta-grid of the verb when used in the middle; the evidence that this argument is also projected into the syntax is however no stronger than the evidence that it is simply present in the theta-grid of the verb.

Turning to the second topic, we saw that one might very normally be tempted to suggest an analysis of the pseudo-middle that would make it similar to the English middle rather than to the French middle. Upon closer examination however, we saw that the similarities with the English middle are rather superficial, and that there are reasons to give the same basic analysis to both the middle and the pseudo-middle, the semantic differences between the two resulting from the role of the surface subject or of the implicit argument being enhanced or downgraded, depending on other features of the sentence.

NOTES

[1] This work has been partially subsidized by Research Grant 410-83-0333 from the Social Scienced and Humanities Research Council of Canada.

[2] Notice for example that **la discussion** is not a possible object of the transitive use of **dérouler:** *"on a déroulé la discussion avec plaisir".

[3] There exists another use that is by and large equivalent:

(i) a. Cette maison s'est vendue **toute seule.**
b. Cette voiture se conduit **toute seule.**

[4] What I have in mind here is the possibility that the pseudo-middle be treated as a simple reflexive construction.

REFERENCES

Demonte, V. 1985. Predication and Passive. Manuscript. Universidad Autónoma de Madrid.

Keyser, S.J. and T. Roeper. 1984. On the Middle and Ergative Constructions in English. Linguistic Inquiry 15.3.381-416.

Hale and S.J. Keyser. 1985. A View from the Middle. Manuscript. M.I.T.

Jaeggli, O. 1984. Passive. University of Southern California.

Lakoff, G. 1977. Linguistic Gestalts. Papers of the 13th Regional Meeting of the Chicago Linguistics Society, 236-286.

Rizzi, L. 1985. Null Objects and the theory of **pro**. Manuscript. M.I.T.

Ruwet, N. 1972. "Les constructions pronominales neutres et moyennes", in **Théorie syntaxique et syntaxe du français**, Editions du Seuil, Paris, 87-125.

Van Oosten, J. 1977. Subjects and Agenhood in English. Papers of the 13th Regional Meeting of the Chicago Linguistics Society, 459-471.

Osvaldo A. Jaeggli

ECP Effects at LF in Spanish

0. Introduction

Much research has been devoted in recent years to providing an account for the contrasts shown in (1) and (2):

(1) a. Who did you say left early?
    b. *Who did you say that left early?

(2) a. Why did you say Bill left?
    b. *Why do you wonder when Bill left?

(2b) is ungrammatical in the reading where why is interpreted as originating in the embedded clause, while (2a) easily allows this construal. These contrasts have been attributied to the Empty Category Principle (cf. Chomsky 1981, 1982, 1986 and references cited there), which states that a nonpronominal empty category must be properly governed.

It is well known that the counterpart to (1b) is grammaticall in Spanish. Cf.:

(3) ¿Quién dijiste que salió temprano?
    Who you+say that left early?

Indeed, omitting the complementizer in these cases leads to ungrammaticality in Spanish. The contrast between Spanish and English is usually accounted for by assuming that in Spanish the trace of a subject is properly governed even in the presence of a lexical complementizer, while this is not the case in English. In English, the trace of a subject is properly governed only in the absence of a lexical complementizer. This difference is generally attributed to the fact that Spanish can freely invert the subject, while this possibility does not exist in English, cf. Jaeggli (1982), Rizzi (1982). On the other hand, Spanish and English pattern alike with respect to adjunct extraction. The trace of an adjunct, in both Spanish and English, can be properly governed only if its antecedent (or its trace) is in the local COMP. Hence, Spanish shows the contrast given in (4), parallel to (2):

(4) a. ¿Porqué dijiste que Juan salió temprano?
Why you+said that Juan left early?

b. ¿*Porqué no sabes cuándo salió Juan?
Why not you+known when left Juan?

Thus, Spanish differs from English with regard to ECP effects involving extraction of a subject over a lexically filled COMP, but it patterns with English with regard to adjunct extraction out of a wh-island.[1]

May (1985) argues that the following sentences also show the effects of the Empty Category Principle.

(5) a. What did everyone buy for Max?
b. Who bought everything for Max?

In (5a), **everyone** can have scope over the wh-element. Thus, it is possible to answer that question by saying: "Bill bought him a tie, Mary bought him a pipe, Peter bought him records, ...". For purposes of exposition, we will call such a reading a **distributed reading**. But this is not possible in (5b). May (1985) attributes this difference to the Empty Category Principle. He points out that if the quantifier **everything** were to undergo QR and adjoin to S in (5b), yielding an LF representation as in (6):

(6) [[Who$_1$][everything$_2$][t$_1$ bought t$_2$ for Max]]

a situation similar to what occurs in (1b) would obtain. The trace in subject position will not be properly governed, violating the ECP. The only possibility available is for **everything** to QR and adjoin to VP, an option which May claims should be allowed on general grounds. In its VP-adjoined position, **everything** cannot be interpreted as having wide scope over the wh-phrase. The situation is different in (5a), as the trace of **what** meets the ECP by being properly governed by **buy**. It need not be antecedent-governed from COMP. Hence, the quantifier in subject position may, and in fact must (for independent reasons), adjoin to S. This allows it to have scope over the wh-phrase in COMP, yielding the desired interpretation.

In short, May (1985) argues that the interpretive possibilities of quantifiers and wh-elements in the cases discussed above show the effects of the ECP applying at the level of Logical Form after the rule of Quantifier Raising has applied.

We saw above that Spanish differs from English with respect to the possibility of subject extraction. In this paper we consider the situation that obtains with the Spanish counterparts to (5). The logic of previous analyses leads us to predict that Spanish should differ from English. The Spanish counterpart to (5b) should allow an interpretation in which the quantifier has scope over the wh-phrase. Traces of subject extraction are assumed to be properly governed in their inverted position in Spanish, irrespective of what relation holds between the trace and the nearest COMP. Adjunction of a quantifier to S should not affect the status of the subject trace with respect to the ECP.

It comes as a bit of a surprise, then, when this expectation is not fulfilled. Cf.:

(7) a. ¿A quién examinó cada doctor?
whom examined every doctor
'Who did every doctor examine?'

b. ¿Quién examinó a cada paciente?
who examined every patient
'Who examined every patient?'

In Spanish, as in English, the (b) sentence does not allow wide scope for the universal quantifier. The (a) sentence does. If this fact is to be attributed to the ECP, as argued by May (1985), previous accounts of subject extraction in Spanish must be modified. More importantly, previous accounts of the Empty Category Principle must also be modified.[2]

In this paper, we undertake these tasks. In section 1, we consider data as in (7) in depth. We show that similar facts hold with a wide variety of verb types: transitives, unergative intransitives, and ergative intransitives. On the other hand, so-called 'psych-movement' verbs show a difference. In section 2, we reconsider the structure of Comp, as well as the central notions involved in the Empty

Category Principle. We adopt a conception of 'government' which incorporates the notion of a 'barrier' as in Chomsky (1986). We then provide an analysis of subject extraction in Spanish in terms of this modified theory of the ECP.

1. Quantifier scope in contexts of wh-extraction in Spanish

Let us begin by considering further data which exhibits the contrast in (7).

The contrast shows up with several different types of transitive verbs, regardless of whether the object is human or not. It also shows up with quantifiers other than **cada** 'every', such as numeral quantifiers. Occasionally we will use numeral quantifiers instead of **cada**.[3] Consider the following sentences.

(8) a. ¿Quién visitó cada ciudad (que María conoce)?
who visited every city that Maria knows
'Who visited every city that Maria knows?'

b. ¿Quién acusó a tres delincuentes?
who accused three delincuents
'Who accused three delincuents?'

(9) a. ¿Qué ciudad visitó cada turista?
which city visited every tourist
'Which city did every tourist visit?'

b. ¿A quién acusaron tres soplones?
whom accused three informers
'Who(m) did three informers accuse?'

The sentences in (8) do not allow a distributed reading for the quantifier in object position. That is, that quantifier cannot have scope over the subject wh-phrase in COMP. On the other hand, the sentences in (9) do allow such readings. The quantifiers in subject position in those sentences may be interpreted with scope over the object wh-phrase in COMP.

The contrast noted above for transitive verbs extends to unergative intransitives. Consider the following sentences:

(10) a. ¿Quién habló en cada conferencia de este ano?
who spoke in every conference of this year
'Who spoke in every conference this year?'

b. ¿Quién durmió en cuatro suites de ese hotel?
who slept in four suites of that hotel
'Who slept in four suites of that hotel?'

(11) a. ¿En qué conferencia habló cada professor?
In which conference spoke every professor
'In which conference did every professor speak?'

b. ¿En qué suite durmió cada invitado?
In which suite slept every guest
'In which suite did every guest sleep?'

Once again, the distributed reading is only available in the sentences in (11). The sentences in (10), which involve the extraction of a wh-subject, do not allow such a reading.

At this point, it is worthwhile to point out that the trace of subject extraction in all of the above cases is generally assumed to be in the same structural position, i.e. VP adjoined position. Following Burzio (1981), Rizzi (1982), and Jaeggli (1982), transitive and unergative intransitive structures share the property of having external subject positions. They differ only in that transitive structures also have internal direct object positions, while this is lacking with intransitive structures. Subject extraction, then, is mediated by the rule of subject inversion, which adjoins the subject to the VP, an option freely available in null subject languages like Spanish and Italian. We might attribute the impossibility of a distributed reading for the cases which involve the extraction of a wh-subject to this common factor. That is, the following generalization would cover all the cases discussed so far.

(12) A quantifier in object position cannot have scope over a wh-subject extracted from an external θ-position.

The intuition behind (12) is that the asymmetry observed so far reflects the fact that in the case where the distributed reading is possible, wh-extraction proceeds from an internal θ-position, while in those instances where the distributed reading is not available, wh-extraction proceeds from an external θ-position. The question that arises at this point, then, is whether ergative constructions, the subjects of which have D-structure VP-internal positions, pattern with the cases discussed so far or not. If they pattern differently, (12) would be further supported. If on the other hand they patterned with the cases discussed so far, (12) will be shown to be an incorrect generalization.

Consider, then, ergative constructions.

(13) a. ¿Quién llegó de cada ciudad que conoce
who arrived from every city that knows

María?
Maria?

'Who arrived from every city that Maria knows?'

b. ¿Qué metal existe en tres planetas del
which metal exists in three planets of+the

sistema solar?
system solar?

'Which metal exists in three planets of the solar system?'

(14) a. ¿De qué ciudad que conoce María llegó
From which city that knows Maria arrived

cada turista?
every tourist?

'From which city that Maria knows did every tourist arrive?'

b. ¿En qué planetas del sistema solar existen
in which planets of+the sysem solar exist

tres metales que nos interesan?
three metals which us interest?

'In which planets of the solar system do metals which interest us exist?'

As in the cases discussed so far, the sentences in (13) which involve the extraction of a wh-subject do not allow a distributed reading for the quantifier inside VP. On the other hand, the sentences in (14) do allow a distributed reading for the quantifiers in subject position. These sentences show that the generalization in (12) is incorrect. The asymmetry does not involve external vs. internal θ-positions. Rather, it appears to be a pure subject/object asymmetry. A quantifier is not allowed to have scope over an extracted wh-subject, regardless of whether the subject originates in an external or in an internal θ-position. The correct empirical generalization, thus, appears to be as in (15):

(15) A quantifier in object position is not allowed to have scope over an extracted wh-subject.

The intuition behind (15) is that the asymmetry in question reflects a difference in the grammatical relation born by the extracted constituent. In instances of subject extraction, wide scope quantification is disallowed.

We can test the empirical validity of (15) by considering sentences which involve wh-extraction of an object. If (15) is correct, distributed readings should be available in these instances. If distributed readings are not available even in these cases, on the other hand, this may indicate that quantifier phrases in object position are never allowed to have scope over a wh-phrase in COMP. Note that all of the cases considered so far where a quantifier phrase has scope over a wh-phrase involve quantifier phrases in subject position.

Consider then the following sentences.

(16) a. ¿A quién entregó María cada regalo?
to whom handed Maria every gift
'To whom did Maria give every gift?'

b. ¿Qué nota le otorgó el profesor a cada
which grade gave the professor to every

estudiante?
student

'What grade did the professor give every student?'

c. ¿Sobre qué habló el profesor con cada
about what spoke the professor with every

estudiante?
student?

'What did the professor speak about with every
student?'

d. ¿Cuándo salió Juan con cada hija del
when went-out Juan with every daughter of

Sr. Martinez?
Mr. Martinez?

'When did Juan go out on a date with every
daughter of Mr. Martinez?

e. ¿Desde cuándo existe el hierro en cada planeta
since when exists the iron in every planet

del sistema solar?
of+the system solar

'Since when does iron exist in every planet of
the solar system?'

These sentences all involve extractions of non-subjects and a quantifier inside VP. As in similar English sentences, all sentences allow wide scope readings, i.e. distributed readings, for the quantifiers inside VP. These sentences give evidence, then, that condition (15) is empirically correct. It is only subject extraction which blocks a wide-scope reading for the quantifiers inside VP. Extraction of non-subjects do not have this property. It is precisely because condition (15) receives this type of confirmation that it appears correct to attribute the noted asymmetry to the ECP. Otherwise we could simply stipulate that wh-elements in COMP always have wide scope over a non-subject quantifier. (In fact, this statement need not be a stipulation, but may be derived from other principles of grammar). But given the evidence presented in (16), such a statement would be empirically unsatisfactory. It would represent an incorrect generalization. Henceforth we will assume that condition (15) represents a first approximation to the facts, and continue to explore the data in this direction.

A counterexample to the generalization in (15) would involve an instance of subject extraction which does allow wide-scope for a quantifier in object position. There is a type of verb which we have yet to consider: so-called 'psych-movement' verbs, verbs such as **gustar** 'to like', **molestar** 'to bother', etc. These verbs take two arguments: a theme argument and an experiencer argument. In Spanish, the theme argument is the grammatical (surface) subject of the sentence. The verb agrees in person and number with this argument. The experiencer argument is expressed as an indirect object. The verb does not agree with the experiencer argument, although this argument may be clitic doubled, and in many cases this doubling is highly preferred. Sentences with these verbs counterexemplify condition (15).

(17) a. ¿Qué actor le gusta a cada chica?
which actor pleases to every girl?
'Which actor does every girl like?'

b. ¿Qué canción de los Beatles le molesta a
which song of the Beatles bothers to

cada adulto que conoces?
every adult that you+know

'Which song by the Beatles bothers every adult that you know?'

(18) a. ¿A quién le gusta cada actor que aparece en
to whom pleases every actor that appears in

ese film?
that film

'Who likes every actor that appears in that film?'

b. ¿A quién le molesta cada canción que
to whom bothers every song that

tocan por la radio?
they+play by the radio

'Who does every song that they play on the radio bother?'

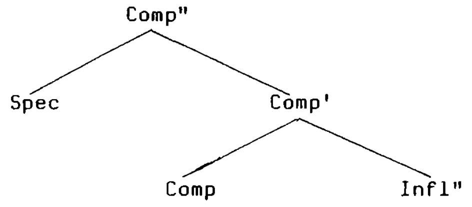
Here the judgments are reversed. The quantifier is allowed to have scope over the wh-phrase in COMP only in the sentences in (17). The quantifier is not allowed to have scope over the wh-phrase in the sentences in (18). Note however, that it is precisely in (17) that the wh-phrase corresponds to an extracted grammatical subject. In (18), the extracted phrase corresponds to the grammatical indirect object of the sentence. And yet, it is only in this case that the quantifier is not allowed to have scope over the extracted wh-phrase. These sentences, then, do not conform to condition (15). In fact, they counterexemplify it in the most direct way. The indirect object in these sentences functions as subjects do in all other sentences. The grammatical subjects of these sentences, on the other hand, function as objects do in the rest of the paradigm.

However, instead of reformulating (15) to take these cases into account at this point, we will leave them as apparent exceptions to this condition. It is only when we have a sufficiently clear idea of how to account for condition (15) that we will bring these examples back into the discussion.

2. The Structure of Comp, Notions of Government, and Barriers

The structure assigned to Comp plays a crucial role in all theories of the Empty Category Principle. In previous analyses, Comp was assumed to be a (maximal) projection sister to S (or Infl"). Chomsky (1986) suggests an alternative analysis. He argues that the structure normally assigned to maximal projections of lexical categories extends to Comp, yielding a structure as in (19):

(19)



All categories, then, lexical and non-lexical, conform to the system provided by X-bar theory. Assume that lexical categories are based on the features [+/-N, +/-V] yielding the categories Noun [+N,-V], Verb [-N,+V], Adjective [+n,+V), and Preposition-Postposition [-N,-V]. Non-lexical categories are Infl and Comp. Assume, following Stowell (1981) that Infl also conforms to X-bar theory. S is then the maximal projection of Infl (=Infl").

The structure in (19) gives two positions under Comp": a head position and a specifier position. It is reasonable to assume that the head position can only be occupied by an X^o category, under Emonds's Structure Preserving Hypothesis. The specifier position may be occupied by a maximal projection.

In non-interrogative structures, the head of Comp position may be occupied by lexical complementizers, such as English **that, for,** and perhaps also **whether** and **if.** It is well-known that in Modern English a wh-phrase cannot coexist with another element in Comp. Consider the following familiar paradigm:

(20) a. I wonder [who (*that) [John met t]].
b. I wonder [what (*whether) [John bought t]].
c. I wonder [when (*if) [John will be here]].

The structure in (19), however, allows for these sentences. Let us assume then, as in much previous work, that sentences as in (20) are disallowed by a version of the doubly-filled Comp filter suitably translated into this framework.[4]

To summarize, we assume that Comp conforms to X-bar theory as in (19), and that a version of the Doubly-Filled Comp Filter regulates whether lexical material may co-occur in the head and specifier positions of Comp.

The central notion in any theory of the Empty Category Principle is the notion of **government.** The concept of government imposes a quite restrictive locality condition.[5] The intuitive content of the notion is that an element governs within its c-command domain those elements which are not protected by a boundary of a particular type. The notion of c-command which appears to be relevant is distinct from the notion of c-command which is operative in binding theory. For the theory of government, the relevant notion appears to ignore non-maximal projections. Accordingly, let us follow Chomsky (1986) (and Aoun & Sportiche 1983) in assuming a notion of **m-command** as in (21):

(21) x m-commands y iff x does not dominate y and every z that dominates x dominates y, where z is restricted to maximal projections.

The notion of government can then be defined as follows (cf. Chomsky 1985):

(22) x governs y iff x m-commands y and there is no z, z a barrier for y, such that z excludes x.

The definition of government in (22) makes reference to three concepts: m-command, exclusion, and barrier. We have already provided a definition for m-command in (21). Before we turn to the other two, consider adjunction structures as in (23):

(23) ...w...[$_z$ x [$_z$... y ...]]

A question that arises concerning such structures is whether the category z dominates x or not. Let us assume, following May (1985) and Chomsky (1986), that in an adjunction structure as in (23), the category z consists of two **segments**, and that an element is dominated by a category consisting of two segments only if it is dominated by both segments of that category. That is, let us assume a principle as in (24):

(24) x is dominated by y only if it is dominated by every segment of y.

Then, in (23) z does not dominate x, since one segment of z does not dominate x. Also, in (23) w and x m-command y, but y does not m-command either of those two categories. Note, in particular, that y does not m-command x, as there is a maximal projection which dominates y but does not dominate x, namely z.

Now let us turn to the notion of exclusion. The intuitive idea is that a category excludes another category if the two are completely independent of each other. Let us assume the following definition:

(25) x excludes y if no segment of x dominates y.

Then, in (23) z excludes w, but it does not exclude x, since x is indeed dominated by a segment of z. Note that in (23), x is neither dominated nor excluded by z. Finally, consider the concept of government, without for the time being taking into consideration the notion of a 'barrier'. For now, let us simply assume that in (23), the category z constitutes a barrier for y. Then, w does not govern y in (23): w does m-command y, but there is a barrier for y which excludes w, namely z. On the other hand, x does govern y in (23). Once again, x m-commands y, and in this case there is no barrier for y which excludes x. In particular, z (which is a barrier for y) does not exclude x, as one of its segments dominates x.

Now let us turn to the concept of a barrier. For some time now, a standard assumption has been that any maximal projection constitutes a barrier to government. But this idea appears to be too strong. Rather, 'barrierhood' appears to be a relative notion. Chomsky (1986) provides the following recursive definition for the concept of a barrier.

(26) x is a barrier for y iff (i) or (ii):
(i) x immediately dominates z, z a blocking category for y, or
(ii) x is a blocking category for y, x = I" (i.e. S)

(27) x is a blocking category for y iff x is not L-marked and x dominates y.

Condition (i) of definition (26) attributes barrierhood (for y) to a category x by inheritance: x can be said to inherit barrierhood (for y) by virtue of dominating a blocking category for y. Condition (ii) of definition (26) establishes that certain blocking categories are inherent barriers, excluding S from this set by stipulation. The definition of blocking category involves the notion of L-marking. This notion is defined as follows:

(28) x L-marks y if and only if x is a lexical category that θ-governs y.

where 'θ-government' is defined as follows:

(29) x θ-governs y iff x is a zero level category that θ-marks y, and x,y are sisters.

L-marking, then, requires θ-government in this system. Finally, 'proper government', the notion directly relevant to the ECP, may be construed as follows: x properly governs y iff x θ-governs y or x antecedent governs y, where antecedent government holds of a link (x,y) of a chain where x governs y.

The theory of government sketched above constitutes a reformulation of several concepts which enter into many of the crucial principles of several sub-components of the theory of grammar. It is natural to expect, then, that it will have a wide range of intricate consequences. We will focus primarily here on some of the consequences which concern the Empty Category Principle, leaving aside many issues of great interest which fall outside the scope of our discussion. Let us begin by reconsidering **that**-trace effects given the theory of government presented above.

Consider a structure as in (30):

(30) [Who did [you say [$_{C''}$ t'that [$_{I''}$ t left]]]]

where **who**, **t'**, and **t** are all coindexed. First let us consider whether **t** is properly governed or not. It is clear that **t** is not θ-governed. However, given the definitions of antecedent government and barrier presented above, **t** is antecedent governed by **t'** in (30). (t', t) constitute a link in a chain, and **t'** governs **t**. I" is a blocking category for **t**, but it is not a barrier: it is not an inherent barrier, and neither does it constitute a barrier by inheritance, since it does not dominate a blocking category containing **t**. Thus, **t** is properly governed in (30): it is antecedent governed by **t'**. What about **t'**? A reasonable assumption is that it is antecedent governed by **who**. In fact, if **that** were not present in (30), the sentence would become acceptable even though the relation that holds between **who** and **t'** would not be affected by the absence of **that**. Note that **who** and **t'** form a chain. Let us assume that they constitute one link in a chain. Does **who** govern **t'**? The I" boundary is not an inherent barrier, though it is a blocking category. But I" does inherit barrierhood from VP, which is also a blocking category under the assumption that VP is not L-marked.[6] Furthermore, VP is a blocking category and a barrier (but see note 6). Finally, C" is not a barrier, as it is L-marked by **say**. Thus, there are two barriers between **who** and **t'** in (30). This technical problem can be resolved if movement to the higher Comp is allowed to proceed via adjunction to VP (see Chomsky 1986 for complete details). (30) should then be revised as follows:

(31) [who did [you [$_{VP}$ t" [$_{VP}$ say [$_{C''}$ t' ...]]]]]

Now **t"** governs **t'**, and **who** governs **t"**. Government holds of the first link since C" is not a barrier, and VP does not exclude **t"**; government holds of the second link, since I" is not an inherent barrier, and VP does not dominate t", hence I" does not inherit barrierhood from VP.

To summarize, both traces in (30) are properly governed. **t** is antecedent governed by **t'**, and **t'** is also antecedent governed as shown in (31). But this is not the correct result. If **that**-trace effects are to be reduced to the ECP, we want at least one of the non-pronominal empty categories involved to fail proper government. Below we suggest how this can be achieved.

3. Head-government in PF

Extraction of a subject wh-phrase in English exhibits **that**-t effects, as we saw above. Adjunct extraction, on the other hand, does not exhibit **that**-trace effects. Thus, the following sentences are all acceptable in an interpretation where **why** modifies the lower clause.

(32) a. Why did Bill say that John left?
b. Why do you think that he got angry?

But adjunct extraction does show ECP effects when the local Comp is occupied by a wh-element, as in (2b) repeated below:

(2) b. *Why do you wonder whether Bill left?

These facts indicate that **that**-trace effects should not be analyzed as involving a failure of antecedent government. We would like to suggest that these facts should be accounted for by making reference to a notion of locality which holds either at S-structure or in the PF component of the grammar, a notion which we call 'head government'.[7] The idea that a part of the Empty Category Principle applies in the PF component is not new. A similar suggestion was made in Jaeggli (1982), and more recently similar ideas have been advanced in Wahl (1985), Stowell (1986), and other sources. Let us assume a notion of head government defined as in (33):

(33) x head governs y iff:
  (i) x is a zero level category,
  (ii) x governs y, and
  (iii) x is a p-governor.

where 'p-governor' is defined as follows:

(34) x is a p-governor iff x is indexed at that level of representation.[8]

The intuitive idea behind the notion of a p-governor is that if a category is unindexed it is not visible for head government at that level. Finally, let us assume that the notion of government involved in (33) is directional. In English, head government holds to the right. Let us now assume that in the PF component all non-pronominal empty categories must be head governed. This is the part of the Empty Category Principle which applies in PF.

Given the definition in (34), consider what may function as a p-governor for empty categories in different syntactic configurations. We assume that all lexical categories (with lexical content) are indexed. Hence, all of them are p-governors for their objects, and more generally for phrases included within their maximal projections. Under the traditional assumption that Infl is an indexed category, Infl is a p-governor. This assumption would also follow if we assume that Infl assigns a θ-role to VP (see note 6 for discussion). Thus, Infl head governs all elements it governs to the right.

Consider subjects next. Subjects in English are not head-governed by Infl, as the directionality requirement is not met. The only category which may head govern a trace in subject position is the head of Comp. But this is only possible if the head of Comp is a p-governor. The head of Comp may qualify as a p-governor only if it is indexed, as it is not a lexical category. In particular, if the head of Comp is occupied by a lexical complementizer, like **that** or **for**, it does not qualify as a head governor, as these complementizers are not indexed. (Note that in a relative clause, **that** may qualify as a head governor if it is assumed that in such constructions this element is indexed. This accords well with the facts, cf. **the man that left early was the criminal**, and recalls several previous analyses.)

We may now analyze **that**-trace effects as a failure of the head government requirement. If the head of Comp contains **that**, it does not qualify as a p-governor for a trace in subject position. Hence, such an empty category cannot meet the head government requirement. Why does the absence of a lexical complementizer in head position make a difference? It must be that this situation allows the head of Comp to function as a p-governor, thus making head government possible.

This effect can be achieved if we assume that the Comp Indexing algorithm of Aoun, Hornstein, and Sportiche (1983) has an analogue within the system developed here. Let us assume that there is a process which applies at S-structure (in English and Spanish) which copies (or checks for agreement between) the index and the WH feature of a phrase in Spec position of Comp and the head of Comp. This rule can be formulated as in (35):

(35) $[_{C''}\ [XP,\ +/-WH_i]\ [_{C'}[_{C}e]\ldots]]]\ ==>$

$[_{C''}\ [XP,\ +/-WH_i]\ [_{C'}[_{C}[+/-WH_i]\ldots]]$

Let us call this process the Spec-Head Feature Rule. That this rule should apply at S-structure in English is motivated by the selection facts of English. Selectional restrictions apply at S-structure in English. Given the standard assumption that selectional requirements are satisfied by the heads of the maximal projections which are selected, a verb which selects a +WH sentential complement will have its selectional requirements met at S-structure only if the head of C" is itself +WH. Similar remarks hold for verbs which select -WH sentential complements.

The Spec-Head Feature rule allows the head of Comp to function as a p-governor for an element in subject position, allowing head government. An acceptable sentence involving subject extraction will then have the following representation at the level at which the head government requirement applies:

(36) Who$_i$ did you say $[_{C''}[t',\ -WH_i]$

$[_{C'}[_{C}[-WH_i]]\ [_{I''}\ t_i\ \ldots]]]$

The trace in subject position is now head governed by the head of Comp, as this category now counts as a p-governor. If **that** occupies the head of Comp position, this is not the case. This follows if we assume that **that** is inserted in the PF component after the Spec-Head Feature rule applies, and that this insertion process destroys the effects of the Spec-Head Feature rule.[9] Note that this would not constitute a non-recoverable deletion, since the effects of **that** insertion are not felt in the LF component of the grammar.

To summarize, the analysis presented above attributes **that**-trace effects to a failure of head government at PF. A trace in subject position can only be head governed by the head of Comp. The head of Comp may act as a head governor only if it contains a WH feature coindexed with the trace in subject position. This situation does not obtain if the head of Comp contains **that** (or **for**, and see note 16): hence **that**-trace effects.

If presence or absence of **that** does not affect the extraction possibilities of adjuncts, it must be the case that adjuncts are head governed at S-structure or PF independently of whatever occupies the head of Comp. Two possibilities come to mind. If adjuncts are generated under VP, they are head governed by V. On the other hand, if adjuncts are generated under Infl', then they are head governed by Infl. In either case, they are head governed independently of whatever occupies the head of Comp. Thus, this analysis predicts that adjuncts should not exhibit **that**-trace effects. We have no evidence bearing on which of the two initial structures is correct for adjuncts. Either one yields correct results given our analysis. Thus, we leave the question open.[10] Finally, for completeness sake we note that traces which are uncontroversially inside VP, like traces of verbal objects, etc., will always be head governed by the verb. Thus, they always meet the head government requirement at PF.

Consider next traces in Spec position of Comp. How do they meet the head government requirement at PF? Recall that C" is not a barrier for a trace in Specifier position of Comp given the definition of barrier in (42). This is so because C" is not a blocking category as it is L-marked by the verb of which it is a complement. It follows, then, that traces in Comp are head governed by the higher verb. Nothing special needs to be added in these cases. A similar result follows for traces in VP-adjoined position, as in (47). Given the definition of head government, Infl head governs these traces. Note that such traces are not dominated by the category they are adjoined to, as they are dominated by only one segment of a category. Thus, the fact that the category VP is a blocking category and a barrier for elements which it dominates does not affect traces which are adjoined to VP.

Let us now turn to consider extraction possibilities in Spanish and the head government requirement in PF. We assume that the set of p-governors of Spanish is identical to the set of p-governors of English. In particular, all lexical categories (with lexical content) are p-governors, Infl is a p-governor, and the head of Comp is a p-governor only if it has undergone the Spec-Head Feature rule. As in English, we assume that this rule applies at S-structure in Spanish. The insertion of the complementizer **que** occurs in the PF component of the grammar.

Spanish tensed sentential complements typically must always be introduced by a lexical complementizer. Absence of a lexical complementizer results in unacceptability. This fact has important consequences for the analysis of extraction presented above. Note that a trace in [NP,S] position (of an embedded clause) in Spanish will never meet the head government requirement in PF, since it cannot be head governed by the head of Comp, as this position will always be occupied by **que** (or another lexical complementizer).

To allow extraction of subjects of embedded clauses, then, we must adopt the analysis that extraction proceeds from the inverted position. In inverted position, the trace of a subject will meet the head government requirement straightforwardly. Traces adjoined to VP, corresponding to subjects of transitive and unergative intransitive verbs, will be head governed by Infl. Traces inside the VP, corresponding to subjects of ergative verbs will be head governed by V. In all cases, then, a trace of a subject will be head governed even though the head of Comp position contains a lexical complementizer. This is the reason why Spanish does not exhibit **that**-trace effects.

The theory presented above nicely accounts for a number of facts which have been attributed to the Empty Category Principle in the recent literature on this matter. It is clear, however, that it does not exhaust the locality requirements which hold in cases of extraction. The Empty Category Principle has been shown to have quite a range of intricate consequences concerning movements which occur in the LF component of the grammar. Since the head government requirement discussed above applies in the PF component, it can have no effect on movements which occur in LF. In the next section, we will consider phenomena which fall under a locality condition which applies in LF.

4. The Empty Category Principle in LF

The notion of proper government invoked by the classical Empty Category Principle (in Chomsky 1981 for example) involved two different types of government: lexical government and antecedent government. The notion of head government introduced above is reminiscent of lexical government. There are important differences, however. Lexical government within the classical ECP involves θ-government in most instances. It is similar to the notion defined in (29) above, repeated below:

(37) x θ-governs y iff x is a zero level category that θ-marks y, and x,y are sisters.

Furthermore, it is restricted to lexical categories. Thus, Infl never counts as a lexical governor. Head government does not involve θ-government, and non-lexical categories, like Infl and Comp are allowed to head govern in certain cases. What about antecedent government? Let us define a notion of a-government as follows:

(38) x a-governs y iff:
(i) x θ-governs y; or
(ii) (x,y) form a link in a chain and x governs y.

This notion is similar, though not identical, to the notion of Identification defined in Jaeggli (1982, p. 149). The first case of a-government involves θ-government: we will call the second type of a-government 'chain government'. There is a considerable evidence that a notion such as a-government contributes further locality restrictions to extraction processes. More specifically, let us assume that at LF the ECP requires that all non-pronominal empty categories be a-governed. We consider some of the evidence that motivates this condition next.

Consider long distance adjunct extraction out of a wh-island, as in (39):

(39) *Why did John know $[_{C''}$ where $[_{I''}$ to go t]]

with t the trace of **why**. This sentence is unacceptable in the interpretation where **why** modifies the lower clause. Following Lasnik and Saito (1984), this fact can be analyzed as failure of a-government. **Why** is never θ-governed, as adjuncts are not arguments which receive a θ-role. Thus, the trace of **why** must meet the a-government requirement via chain government. The unacceptability of (39) follows, then, from the lack of chain government. The trace of **why** is in a chain with its antecedent, but it is not governed by its antecedent, since C" is a barrier for t. C" inherits barrierhood from I", which is a blocking category for t.

Notice that the unacceptability of (39) also provides direct evidence that there must be some other component to the ECP aside from head government in PF. The trace of **why** is head governed in PF in (39). The unacceptability of (39) shows that this cannot be enough. Thus, we have evidence for the notion of a-government holding in LF.

Consider next subject extraction. Subjects are also never θ-governed (in English), as they are not directly θ-marked by a zero level category. They must also meet the a-government requirement via chain government. Once again, if they are extracted out of a wh-island, the structure is ruled out because the trace will not meet a-government. An example is given in (40):

(40) *Who did you wonder $[_{C''}$ what $[_{I''}$ t bought]]

where t is the trace of subject extraction. This trace is not a-governed, as it is not θ-governed, and neither is it chain governed for the same reasons as the adjunct trace is not chain governed in (39). In both cases the correct results follow from the a-government requirement.

Let us return now to consider the facts discussed at the beginning of this paper concerning the scopal properties of certain quantified expressions in contexts of wh-extraction. Recall that the English sentences discussed in May (1985) fall under the following generalization:

(41) In contexts of subject extraction, a quantifier in complement position may not have wide scope over the subject wh-phrase.

The basic fact to be accounted for concerns the difference in scope properties in the sentences in (5), repeated below:

(5) a. What did everyone buy for Max?
b. What did John give to everyone?
c. Who bought everything for Max?

The quantifier in (5a,b) can be interpreted as having scope over the wh-element in Comp, but this interpretation is not available in (5c). As in May (1985), we assume that a quantifier is allowed to have scope over a wh-element in Comp only if it is adjoined by QR to I". The sentences in (5) indicate, then, that the quantifier is allowed to adjoin to I" in (5a,b), but not in (5c). Do we have an account for this restriction given the theory outlined above? Consider the LP representation of (5c) under the unavailable reading.

(42) $[_{C''}\text{who}_i\ [_{C'}[_{C}[+\text{WH}_i]]\ [_{I''}\ \text{everything}_j$
$[_{I''}\ t_i\ \text{bought}\ t_j]]]]$

The trace of **who** is a-governed by **who** via chain-government. It forms a link of a chain with **who**, and **who** governs it since I", though a blocking category, is not a barrier. At PF it is head governed by the head of Comp, which has been given an index and the +WH feature by the Spec-Head Feature rule. This representation, then violates none of the locality requirements we have considered so far. Something else must be involved if these contrasts are to follow from the Empty Category Principle. Let us assume, then, that aside from a-government, a form of head government must also hold at LF. We can account for the ungrammaticality of the representation in (69) if we assume that head government does not hold across a structure of adjunction.[11] Thus, in a structure as in (43):

(43) $x \ldots [_{z}\ w\ [_{z}\ y \ldots]]$

x does not head govern y. Let us make this assumption and explore some of its consequences concerning the cases in (5).

Suppose, then, that head government must also hold in LF, and that head government does not hold in structures like (43). It follows that (42) will be ungrammatical since the head government requirement will not be met for the trace in subject position. On the other hand, in (5a) and (5b), the traces of wh-movement will be head governed even if the quantifier adjoins to S. Since those traces are not head governed from Comp, adjunction to S will have no effect. Thus, this possibility will be allowed, yielding a wide scope interpretation for the quantifiers. The central cases discussed by May, then, follow naturally given this assumption.

So far we have considered the effects of certain LF locality requirements on structures of extraction in English. Next we will consider the corresponding structures of extraction in Spanish.

Spanish and English differ markedly with respect to long distance subject and object extraction out of a wh-island. In English, long distance subject extraction out of a wh-island yields a strong case of unacceptability, while object extraction out of a wh-island is at times only marginal (at worse). The sentences in (44) illustrate this.

(44) a. **Who$_i$ do you wonder what$_j$ t_i bought t_j?

b. ??What$_i$ do you wonder where$_j$ to buy t_i t_j?

(44a) involves an ECP violation, while there is no ECP violation in (44b), only a Subjacency violation. As is well known, ECP violations tend to be much more unacceptable than Subjacency violations. In Spanish the facts are quite different.

In Torrego's (1984) work on wh-movement in Spanish, it is pointed out that long distance subject extraction out of a wh-island is perfectly acceptable, while long distance object extraction out of a wh-island is rather marginal.[12] Consider the following two paradigmatic examples:

(45) a. ¿Quién no sabes qué compró?
'Who don't you know what bought?'

b. *¿Qué no sabes quién compró?
'What don't you know who bought?'

This situation is the exact opposite of what obtains in English. How is this difference to be accounted for? Let us begin by considering the unacceptable (45b).

It is well-known that in Spanish in a sentence containing an argument wh-element in Comp, the verb must prepose.[13] Following the theory of Comp presented in section 2, let us assume that the verb preposes to the head of Comp. Then, the structure at S-structure of the embedded C" in (45b) is:

(46) $[\text{qué}_j \ldots [_{C''}\text{quién}_i\ [_{C'}[_{C}\text{compró}]$

$[_{I''}\ \text{pro Infl}\ [_{VP}[_{VP}[_{V}e]\ t_j]\ t_i]]]]]$

where **pro** is an expletive pure pronominal. We continue to assume that subject extraction proceeds from post-verbal position as in section 4 above. As in Torrego (1984), we assume that the verbal trace left by V-movement is not a θ-governor. The trace of V-movement does count as a head governor, though, as the verbal trace bears the index of its antecedent, and the antecedent is coindexed with the object position by virtue of θ-role assignment. Head-government is clearly required for simple cases of object extraction. But if θ-government is not an option in such structures, the trace of object extraction must be a-governed via chain government.[14] But in (46), the trace in object position is not chain governed. Its antecedent is too far away to chain govern it. There are two barriers intervening between **que** and its trace: C" and VP. Note that if extraction proceeded via adjunction to the lower VP, still there would be a trace which failed to be a-governed, namely the trace in VP adjoined position. This case, then, is ruled ungrammatical due to the fact that the trace of the object is not a-governed.

Consider next long distance subject extraction. This case contrasts with long distance object extraction. Let us take a closer look at the structure in question. (45a) has the following structure as S-structure:

(47) $[\text{Quién}_i \ldots [_{C''}\ \text{qué}_j\ [_{C'}[_{C}\text{compró}]$

$[_{I''}\ \text{pro Infl}\ [_{VP}[_{VP}[_{V}e]\ t_j]\ t_i]]]]]$

t_j is head governed by the empty verb, and chain governed by its antecedent in the local Comp. What about the trace of the subject? It is head governed by Infl, as argued in section 4 above. How is it a-governed? It clearly is not chain-governed by its antecedent, since its antecedent (in all possible cases) is too far away. We would like to suggest that this trace is θ-governed by Infl, and meets the a-government requirement this way. Since it is θ-governed, it can be extracted long distance without any problem. Let us see how θ-government by Infl obtains in these situations.

Consider a structure of subject inversion.

(48) $[_{I''}$ pro $[_{I'}$ Infl $[_{VP}$ VP NP]]]

where **pro** is an expletive pure pronominal. How is the NP in inverted position θ-marked? To answer this question we need to address the more general issue of (subject) θ-marking. Let us say that a category x may θ-mark a category y iff x,y are sisters, where the 'sister' relation is defined as in (49):

(49) x and y are sisters if they are dominated by the same lexical projections and they exclude each other.

Under this definition of sisterhood, adapted from the definition given in Chomsky (1986), VP and NP are not sisters in (48), since VP does not exclude NP. Then, VP is not allowed to θ-mark NP. Notice, however, that Infl is a sister to NP, as Infl and NP exclude each other and Infl and VP are sisters, too. Let us assume, then, that θ-marking of the subject NP in inverted structures is done as follows: the VP, which contributes the relevant θ-role, θ-marks Infl. Infl, in turn, θ-marks the inverted subject. Infl must get rid of the θ-role as it is not an argument. It does this by assigning it to the inverted NP. In an inverted structure as in (48), then, Infl θ-marks the subject. But then, Infl θ-governs the inverted subject position, since Infl is a zero level category that θ-marks that position, and it is a sister to that position.

This analysis claims that a subject trace is θ-governed by Infl. Long distance extraction, then, is not a problem, as the trace need not rely on an antecedent in the local Comp to meet the a-government requirement. Notice, however, that a trace of a subject can be a-governed by Infl only if the subject has been post-posed. We do not want to allow a trace in [NP,S] position to be θ-governed by Infl in Spanish, as this would allow extraction from that position in LF. It can be shown that extraction from that position is not allowed for wh-elements.[15] Consider the following sentences, instances of multiple interrogation which contain wh-elements in situ.

(50) a. ¿Quién dijiste que compró qué?
who you-said that bought what?
'Who did you say bought what?'

b. ¿Qué dijiste que compró quién?
what you-said that bought who
'What did you say that who bought?'

c. ¿*Qué dijiste que quién compró?
what you-said that who bought
'What did you say that who bought?'

d. ¿Qué dijiste que Mario compró?
what you-said that Mario bought
'What did you say that Mario bought?'

In (50a), the wh-element in situ is in object position. It is wh-raised to the matrix Comp, occupied by **quién** at LF; and its trace meets both the a-government requirement (since it is θ-governed by the verb) and the head governed requirement. In sentence (85b), the trace of **quién** left by wh-raising at LF also meets both requirements, it is both a-governed and head governed by Infl. Consider now (50c). Note that this sentence is not ruled out because Verb preposing did not apply. Verb preposing is not obligatory in these structures, as sentence (50d) shows, since the embedded Comp need not contain a trace of a wh-element, as the object wh-element in the matrix Comp can move up there in one step without violating Subjacency given that only C" is a bounding node in Spanish. The contrast between (50c) and (50d) shows quite clearly that the ungrammaticality of (50c) should be attributed to the fact that there is a wh-element in [NP,S] position. The wh-element in situ is raised to the higher Comp by wh-raising. The trace of this rule violates the ECP under the assumption that it is neither head nor a-governed. If Infl were allowed to head and a-govern the trace of **quien** in (83c), the sentence would be ruled grammatical. Thus, we must ensure that Infl only head and a-governs traces of subjects which have been extracted from post-verbal position in Spanish. As regards head government, this requirement follows from the idea that the notion of government involved is directional. If head government always occurs to the right in Spanish, Infl will never head govern a trace in [NP,S] position. But a similar result does not yet follow for θ-government by Infl. We cannot claim that θ-government only holds to the right in Spanish, at least not if θ-government is defined in terms of sisterhood and θ-marking. Sisterhood, as defined above, is not a directional concept, and neither is θ-marking, especially not in the case of subjects. Although objects are always θ-marked to the right in Spanish, this is not true precisely for subjects. Subjects must be capable of being θ-marked in [NP,S] position, otherwise even simple sentences with pre-verbal subjects will be ruled out by the θ-Criterion. However, there is a crucial difference between subjects which are θ-marked in inverted position. A subject in [NP,S] position is a sister to VP, as it is excluded by VP. Hence, it may be θ-marked by VP, just as it is in English. It is only if the subject has been inverted and adjoined to VP that it cannot be θ-marked by VP, as it not excluded by this category. Let us assume then, that when a subject is in [NP,S] position, it is always θ-marked by VP, and never by Infl. We can state this as the following principle:

(51) A zero level category can only θ-mark to the right
(in Spanish).

The category VP is allowed to θ-mark to the left as it is not a zero level category. Infl, on the other hand, is not allowed this option given principle (51). It follows, now, that when a subject is in pre-verbal position, it will always be θ-marked by VP, and when it is in inverted position, it will always be θ-marked by Infl. We assume that θ-marking must occur by S-structure, hence these results are sensitive to the position occupied by a subject at S-structure.

Note that this account solves a long-standing problem concerning the derivations which allow subject extraction at LF. We know that subject extraction at LF should not be allowed from [NP,S] position. But it has always remained unclear what prevents a derivation where subject extraction

at LF proceeds via VP adjunction first. The theory outlined above accounts for this impossibility, as even in these cases the trace of the subject in adjoined position will fail to be θ-governed by Infl. Hence an ECP violation will result even in these cases.

Returning now to (50c), it is ruled out because the trace of subject extraction will not be a-governed, as it is not θ-governed by Infl. Note that a subject in [NP,S] position is never θ-governed by VP, even though VP is a sister to the subject NP and it θ-marks this category, because VP is not a zero level category, and only zero level categories are θ-governors. (See definition (45) above).

To summarize, the theory sketched above allows for long distance subject extraction in cases like (45a) in Spanish because the trace of subject extraction is head and a-governed by Infl. Similar instances of LF extraction, as in (50b) are also allowed, while improper cases of LF extraction, as in (50c) are correctly ruled out. Note that this analysis captures the basic insight of Rizzi's original analysis of subject extraction, without, however, implicating the verb. But the analysis we propose still captures the core of Rizzi's insight, namely that subject extraction in Italian (or Spanish, for our purposes) is allowed only if it proceeds from post-verbal position. Thus, our analysis extends with full generality to the other cases discussed in Rizzi (1982) (e.g. the **nessuno** scope facts).

While subjects extraction out of a wh-island is quite acceptable in Spanish, contrasting sharply with English, the two languages behave alike with respect to adjunct extraction of a wh-island. The English case was illustrated above in sentence (39), which we repeat below in (52):

(52) *Why$_i$ did John know [$_{C''}$ where [$_{I''}$ to go t$_i$]]?

The trace of **why** does not meet the a-government requirement, as it is not chain-governed from the local Comp and neither is it θ-governed. Analogous cases in Spanish are also unacceptable (in the intended interpretation). Cf.:

(53) *¿Porqué$_i$ no sabes [qué [comprar t$_i$]]
why neg you*know what to+buy
'Why don't you know what to buy?'

The analysis for Spanish is the same as the one for English.

(54) a. ¿Quién vio a cada estudiante?
who saw every student
'Who saw every student?'

b. $[_{C''}$Quién$_i$[vio $[_{I''}$ cada estudiante$_j$

$[_{I''}$pro Infl $[_{VP}[_{VP}[_{V}e]\ t_j]\ t_i]]]]$

We are concerned with the status of t_i. The analysis sketched above for cases of long subject extraction involved the assumption that this trace is head and θ-governed by Infl, a possibility which exists if the subject is inverted and extracted from VP adjoined position. But if this is correct, this structure is well-formed and the quantifier phrase **cada estudiante** should be interpretable with wide scope over the subject wh-phrase **quién**. However, we saw that this interpretation is not available in Spanish any more than it is in English. Thus, something must be added to our story in order to capture these facts.

We propose an analysis which incorporates a suggestion made in Chomsky (1985) concerning expletive elements. For reasons which are independent of the issues that concerns us here, Chomsky (1985, p. 250) suggests that "expletives are not permitted to appear in LF-representation. That is, at LF ... only arguments or their traces appear in A-positions, a rather natural requirement given the role of LF as representing the contribution of the language to semantic interpretation." Chomsky suggests that an expletive x "can be eliminated in accordance with the condition on recoverability of deletion only if x is replaced by a coindexed element y, hence by movement of y to the position occupied by x, forming a chain (y,e)". Let us call this process **expletive substitution.** This process, as other movement processes, leaves a trace subject to the requirements of the Projection Principle (and perhaps other principles of grammar). Since the trace of expletive substitution plays no role in the following analysis, we ignore it throughout in the interest of simplicity.

Assuming a process of expletive substitution, the LF representation of (54a) is no longer (54b) but rather (55):

(55) $[_{C''}$Quién$_i$[vio $[_{I''}$ cada estudiante$_j$

$[_{I''}t_i$ Infl $[_{VP}[_{V}e]\ t_j]]]]$

This representation is much closer to the LF representation of similar sentences in English. In a sense, expletive substitution has the effect of making Spanish and English very similar at the level of LF, at least with respect to the position of the subject. Notice, in particular, that the trace in subject position in (55) does not meet the head government requirement, just like traces in similar structures in English do not meet this requirement either. The adjunction of the quantifier phrase to I" blocks head government. Thus, this structure is ill-formed with respect to the head government requirement at LF, and the unavailable interpretation is correctly excluded. The structure in (55), then, yields the correct result with respect to the scope of the quantifier in instances of subject extraction.

Object extraction, on the other hand, is unaffected by expletive substitution or adjunction of the quantifier to I". We saw that the trace of an object is head governed by the (empty) verb, and chain-governed from the local Comp. As only head government, but not chain government, is blocked by adjunction, no ill effects follow from quantifier adjunction to I" in instances of object extraction. This is exactly as required by the facts. In instances of object extraction, quantifiers are allowed to have scope over the wh-element in Comp. (We return to psychological predicates and indirect objects below.)

However, expletive substitution raises a problem which we must address at this point. Consider the grammatical LF representation for (54a) after expletive substitution:

(56) [$_{C"}$Quién$_i$[vio [$_{I"}$ t$_i$ Infl

[$_{VP}$cada estudiante$_j$[$_{VP}$[$_V$e] t$_j$]]]]

Here **cada estudiante** does not have scope over **quién**, as it does not c-command (or govern) **quién**. But the question that arises at this point is how is **t$_i$** head governed in this structure? Recall that head government at LF requires government by a coindexed zero level category. What category is fulfilling this task in (56)? Note that the verb is in the head of Comp position in (56). Let us assume that the verb arrives into that position by first adjoining to Infl and then being raised with Infl to the head of Comp position. Several recent studies have suggested that at LF there is a systematic relation holding between Infl and the head of Comp position. Let us interpret this relation as a movement process, which raises Infl to the head of Comp. This process may occur in the syntax (in interrogatives) or at LF. A verb may move to the head of Comp position only if it attaches to Infl. Assuming this process then, the correct LF-representation for (54a) is not (54) but rather (57):

(57) $[_{C''}\text{Quién}_i[[_I[\text{Infl}]\text{vio}]\ [_{I''}\ t_i\ [_I e]$

$[_{VP}\text{cada estudiante}_j[_{VP}[_V e]\ t_j]]]]$

Now t_i is head and θ-governed by Infl in the head of Comp position once again. If a quantifier is adjoined to I", though, the head government relation will be disturbed, and the sentence will be out as before.

Notice that the movement process which raises Infl to the head of Comp position may apply either in the syntax or at LF. It applies in the syntax in instances when a verb occupies the head of Comp position at S-structure. Subject-auxiliary inversion in English, and verb preposing in Spanish, can be seen as a reflexes of this movement process applying in the syntax. Interesting questions arise concerning the obligatoriness or optionality of this movement process in different structures (and in different languages). Thus, in many languages the inversion process in only required in matrix interrogatives, while it is disallowed in embedded interrogatives. This is the situation in English, for example. In other languages, the inversion process is obligatory in both matrix and embedded interrogative contexts depending on what occupies the Specifier of Comp position. This is the situation that obtains in Spanish. An analysis of what motivates the rule, or what is responsible for its obligatoriness or optionality in different contexts lies beyond the scope of this paper. For the purposes of our concerns here, it is sufficient to note the effects of this rule.

The rule must be allowed to apply at LF in instances when the verb does not prepose in Spanish. As mentioned in note 29, long distance subject extraction out of a wh-island is possible even in instances where the verb has not been preposed to the head of Comp position. Since after expletive substitution the trace of the subject will be in [NP,S] position even in Spanish, it must be head and a-governed in that position given the sentences' acceptability. Thus, we conclude that Infl raising applies at LF in these cases, even if the verb has not been raised. Note that Infl head and θ-governs the subject in Spanish only if the subject is syntactically extracted out of VP adjoined position, and it does this by virtue of the fact that it θ-marks the subject. θ-marking by a zero level category is equivalent to θ-government, and since we assume that coindexing is involved, it also results in head government. Note that the coindexing relation is licensed by θ-marking. It cannot be the case that the automatic coindexing that is usually assumed to exist between the NP in [NP,S] position and Infl is sufficient for head government in these cases. This seems natural as this automatic coindexing does not express any semantically meaningful relation and may be considered invisible at LF.

Expletive substitution is a process which applies in all sentences containing expletive elements in an argument position. In Spanish it applies in all sentences with inverted subjects in VP adjoined position, and it also applies in sentences with ergative verbs. In the latter instance, movement originates from a θ-position internal to VP, with a resulting chain between [NP,S] position and an internal position. Let us assume, now, that after expletive substitution, the variable is in [NP,S] position. This element, then, must meet the ECP requirements of head government and a-government. As before, adjunction will make head government impossible. Thus, even sentences with ergative intransitive verbs will show the effect observed above for transitive sentences, i.e. in the context of subject extraction a quantifier is not allowed to have scope over the wh-element in Comp. This analysis, then, naturally accounts for the facts regarding quantifier scope in contexts of subject extraction with transitive, unergative intransitive, and ergative verbs discussed in section 1 above.

The only cases left to discuss are the sentences which contain "psychological" verbs, e.g. those in (17) and (18), repeated below:

(17) a. ¿Qué actor le gusta a cada chica?
which actor pleases to every girl
'Which actor does every girl like?'

b. ¿Qué canción de los Beatles le molesta a
which song of the Beatles bothers to

cada adulto que conoces?
every adult that you+know

'Which song by the Beatles bothers every adult
that you know?

(18) a. ¿A quién le gusta cada actor que aparece en
to whom pleases every actor that appears in

ese film?
that film

'Who likes every actor that appears in that
film?'

b. ¿A quién le molesta cada canción que
to whom bothers every song that

tocan por la radio?
they+play by the radio

'Who does every song that they play on the radio
bother?'

In these sentences extraction of the indirect object causes the same effect as extraction of the subject in other sentences. Thus, in (18) the quantified expression cannot be interpreted as having scope over the indirect object wh-phrase in Comp. On the other hand, in (17), where the grammatical subject of the sentence has been extracted, there is no difficulty in assigning wide scope to the quantified indirect object.

These facts follow naturally from our analysis if we assume that at LF, the grammatical indirect object of the sentence is raised to [NP,S] position and becomes the 'logical' subject of the sentence. In other words, expletive substitution operates in these sentences in such a way that the grammatical indirect object of the sentence is substituted into the position occupied by the expletive element at S-structure. Let us assume that this process applies at LF to psychological predicates. Then, traces left by extraction of the indirect object behave like subject extraction traces at LF. The LF representation of sentence (18a) with wide scope for the quantified NP then would be as in (58):

(58) $[_{C''}$ a quién$_i$ $[_{C'}[_{C}$Infl+le gusta] $[_{I''}[$cada

actor que aparece en ese fil$]_j[_{I''}$ t_i $[_{I}3]$

$[_{VP}[_{V}e]$ t_j $t_i]]]]]$

It is not surprising, then, that adjunction of a quantified phrase to S should result in an ECP violation. Once again, the head government requirement at LF will not be satisfied for the trace in [NP,S] position. Thus, these sentences are out with the incorrect interpretation for the same reason as the other sentences discussed above. Our analysis accounts for the impossibility of wide scope quantification in all these sentences in a unitary way. The difference between them is reduced to the way in which expletive substitution operates with different verb classes. In most instances, this process substitutes the grammatical subject of the sentence into [NP,S] position. With psychological predicates, the substitution mechanism raises the grammatical indirect object of the sentence. This difference accounts for the difference noted in section 1 concerning wide scope quantification in contexts of wh-extraction.

NOTES

* I am indebted for helpful comments and discussion to Joseph Aoun, Mamoru Saito, Tim Stowell, Nina Hyams, Esther Torrego, and John Hawkins.

[1] Spanish and English show other ECP related differences which we ignore in this paper, in particular, with respect to so-called Superiority effects. In English, the contrast in (i) can be made to follow from the ECP.

(i) a. Who saw what?
b. *What did who see?

In Spanish, the sentence corresponding to (ib) is grammatical, as is predicted by several ECP analyses of these violations. Cf.

(ii) ¿Qué vio quién?
what saw who

However, not all Superiority Condition violations are analyzable under the ECP. Ungrammatical sentences as in (i):

(iii) *What did you tell who(m) that Peter bought?

do not violate the ECP, yet they fall under the descriptive generalization of the Superiority Condition. See Pesetsky (1982) for discussion. For extensive discussion of the ECP in Spanish, see Torrego (1984). We return to some of the facts investigated in that work below.

[2] For a discussion of why the standard version of the ECP, as in Chomsky (1981, 1982) and references cited there, fails to account for these facts, see Jaeggli (1987). Cf. also May (1986).

[3] For reasons which we do not fully understand, some speakers of Spanish do not accept the quantifier **cada** in any position but subject position. Obviously, this is not the case in my dialect of Spanish. For such speakers, nevertheless, the contrasts observed in the text hold with complete regularity with other quantifiers which they do accept in object position, such as numeral quantifiers.

[4] The exercise is not trivial. Note that if a verbal element occupies the head of Comp position, such as an auxiliary or modal verb in English or a verb in Spanish, the filter in question must not rule the sentence out. Thus, the filter must be made sensitive to the contents of the head of Comp. We leave the issue open, as it falls outside the scope of this paper.

[5] The following material concerning the notion of government and other auxiliary notions follows Chomsky (1986) quite closely. We point out where we diverge from the definitions given in that paper.

[6] This assumption might be considered questionable. It has been suggested that Infl bears a relation with VP which is similar to θ-marking. Cf. Fabb (1983). If this is correct, then VP would be L-marked and would not be a blocking category. Then, I" would not inherit barrierhood from VP, and an element in Comp would directly govern another element in a locally subordinate Comp. This idea

raises a number of possibilities and problems which we leave open here.

[7] Here we depart from the analysis given in Chomsky (1986). There it is suggested that the **that**-trace effect follows from the ECP if it is assumed that a lexical complementizer has the effect of making C' into a barrier. This is accomplished via a minimality condition on government. If C' is a barrier for t in a structure as in (30), then it cannot be antecedent governed by t". Hence an ECP violation.

[8] A similar assumption is made in Wahl (1985). The analysis we develop proceeds along slightly different lines, however.

[9] An account along these lines would also explain the impossibility of Aux-inversion in cases of subject extraction if it is assumed that if the head of Comp contains an auxiliary it cannot function an a p-governor. The extension seems straightforward, though we have not considered it in detail.

[10] For different judgments, see Wahl (1985), where it is claimed that adjunct extraction does exhibit **that**-trace effects. If there is such a dialect split, which we are skeptical about, a straightforward revision of the analysis in the text will accomodate those facts.

[11] This idea recalls a suggestion (attributed ty Kyle Johnson) in Chomsky (1986) to the effect that adjunction blocks θ-marking, although the principle we need is not reducible to this suggestion.

[12] We restrict ourselves to a discussion of those cases where the verb has been preposed, and leave aside those instances in which only part of a periphrastic verbal expression is fronted.

[13] In fact, Torrego (1984) provides evidence for a more general statement; namely, that if Comp contains an argument wh-element or its trace, the verb of that clause must prepose. The statement given in the text suffices for our purposes.

[14] Note that the V in Comp must also not count as a θ-governor for the trace in object position. This may be due to the fact that VP is a barrier for the trace in object position or because the V in Comp is not a sister to the trace. In either case, the V in Comp does not count as a θ-governor for the trace in object position.

[15] A similar claim was made for negative quantifiers in Spanish in Jaeggli (1982) (and see Rizzi (1982) for the corresponding claim for Italian). However, there appears to be considerable dialectal variation concerning those facts, thus we prefer to consider wh-in-situ here, concerning which there is no dialectal variation, as far as we know.

[16] A similar analysis covers the Superiority Condition violations which are analyzable under the ECP.

REFERENCES

Aoun, J., N. Hornstein, and D. Sportiche. 1981. "Some Aspects of Wide Scope Quantification," **Journal of Linguistic Research** 1, 69-95.

Belletti, A. and L. Rizzi. 1983. "The Syntax of 'ne': Some Theoretical Implications". **The Linguistic Review** 1.2.

Chomsky, N. 1973. "Conditions on Transformations," in S. Anderson and P. Kiparsky (eds.) **A Festschrift for Morris Halle**. Holt, Rinehart and Winston. New York. NY.

Chomsky, N. 1981. **Lectures on Government and Binding**, Foris Publications.

Chomsky, N. 1982. **Some Concepts and Consequences of the Theory of Government and Binding**. MIT Press.

Chomsky, N. 1985. **Knowledge of Language**. Praeger.

Chomsky, N. 1986. **Barriers**. MIT Press.

Jaeggli, O. 1982. **Topics in Romance Syntax**. Foris Publications.

May, R. 1985. **Logical Form: Its Structure and Derivation**. MIT Press.

Pesetsky, D. 1982. **Paths and Categories**. Unpublished Ph.D. Dissertation. MIT. Cambridge, MA.

Rizzi, L. 1982. **Issues in Italian Syntax**. Foris Publications.

Stowell, T. 1981. **Origins of Phrase Structure**. Unpublished Ph.D. Dissertation. MIT. Cambridge, MA.

Stowell, T. 1986. "Null Antecedents and Proper Government," in S. Berman et al. (eds.) **Proceedings of NELS XVI**, GLSA, University of Massachusetts, Amherst, MA.

Torrego, E. 1984. "On Inversion in Spanish and Some of its Effects," **Linguistic Inquiry**, 15, 103-130.

Wahl, A. 1985. "Two Types of Locality". Unpublished paper, University of Maryland.

Richard D. Janda and Terrell A. Morgan

El acentó dislocadó – pues cantadó – castellanó: On Explaining Stress-Shift in Song-Texts from Spanish (and Certain Other Romance Languages)

To adepts in traditional Spanish poetic-metrics and to those who know a substantial number of songs with Spanish texts, it may be no secret that, in Hispanic music, primary word-stress is often shifted from normally-stressed to normally-unstressed syllables.[1] Thus, for example, as early as 1887, it was noted by Eduardo de la Barra that the expected paroxytonic stress of the word **cóntra** 'against' is 'intolerable' in the rhythmic context of the fourth line of the chorus of the Chilean national anthem and is therefore altered to oxytonic **contrá**. That is, as shown in (1), stressing **cóntra** in the usual way would violate the anapestic trimeter which otherwise uniformly characterizes the verses under consideration, whereas the unusual stress-pattern **contrá** fits in perfectly with its metrical context - contrast especially /o⁀**el así/lo cóntra/la⁀opresión**/ 'or the refuge from oppression' with /o⁀**el así/lo contrá/la⁀opresión**/.

(1) Dul-ce Pá- tria, re-cí- be los vó- tós
Con que Chí- le⁀en tus á- ras ju-ró
Que,⁀o la túm- ba se-rás de los lí- brés
O⁀el a-sí- lo con-**trá** la⁀o-pre-sión.

De la Barra therefore labeled this unusual stress 'anti-rhythmic', saying (p.19) 'el acento...de **cóntra**...es insoportable, i nos obliga a pronunciar **contrá**, para conservar la cadencia rítmica. Yo llamaria [a este acento] antirítmico.'

Nevertheless, a wide-ranging survey of the Hispanic (and even the general Romance) linguistic literature on accentual phenomena brings to light only one other brief mention of musically-conditioned stress-shift in Spanish songs - a short footnote in the Real Academia's (1973:66, n. 7) new grammatical sketch, in which it is stated that

> 'el ritmo es un elemento tan esencial y de primer orden en la palabra hablada como en la música. Lo prueba el hecho de que en el canto, a pesar de la regularidad de los intervalos rítmicos y de la libertad melódica de la música, extrañas a la elocución, lo único que percibe el oído como anomalía es el desajuste entre el ritmo de la tonada y el de la letra, provocado por una distorsión del acento de intensidad....'

Beyond this, there has apparently never been any extensive discussion of the facts of Spanish musical stress-shift, much less an explanation for it. In short, neither in de la Barra's **Elementos de métrica castellana** nor in the Academia's **Esbozo de una nueva gramática de la lengua española** nor in the entire linguistic output of the nearly one hundred years that separate them has there appeared any study treating the implications of this stress-shift for Spanish phonology and/or establishing how and why Spanish differs from a language like, say, English - where musically-conditioned deaccentuation of normally-stressed syllables and accentuation of normally-unstressed syllables is virtually unattested. Somehow, the relevant facts and possible explanations for them seem to have fallen between the chairs of linguistics and music.

In the present paper, however, we begin by setting out the factual parameters of stress-shift in song-texts from Spanish. Next, we propose and evaluate various hypotheses intended to account for the frequency of the phenomenon at issue in Spanish (and in certain other Romance languages, plus at least one non-Romance one) vs. its rarity in English (and various other Romance languages). Finally, we employ these findings as evidence in favor of a number of theoretical claims that have previously been made about Spanish stress on the basis of non-musical considerations. In presenting examples, we have employed the following notational conventions: orthographic accent-diacritics are generally omitted; instead, acute accents mark the actual stresses which constitute the musical rhythm of song-texts. And, most importantly, musically-conditioned stresses which differ from the expected stress in spoken utterances are highlighted in boldface - or, in one case, by capitalization, with orthographic accents then retained. Lastly, cases of synalepha - that is, vowel-coalescence across a word-boundary - are indicated by ligatures. To avoid symbol clutter, these are superscripted rather than subscripted, as is more customary.

Our first concern is to demonstrate that the musical shift of stress in question here is not restricted to merely an insignificant and exceptional subset of Spanish songs, and also that it is genuinely different from possible stress-shifts in any style of spoken (that is, nonsung) Spanish. To this end, we have collected examples documenting the pervasiveness of musical stress-shift in Spanish across time, space, and vocal genres. An extensive list of such representative examples is given in (2). First, as in (2.i), we have collected illustrations from every century (A.D.) since the fifteenth. Second, as in (2.ii), we have assembled instances from twenty-two countries where Spanish is spoken - including the U.S. and the Judeo-Spanish colonies, in addition to every country where Spanish is the national language. And, third, as in (2.iii), we have amassed samples from a variety of musical styles, ranging from pop-rock to opera to flamenco to religious hymns to folksongs to classical art-songs to Christmas carols and more.

(2) **i. Across Time**

a. **Fifteenth century ['Cucú', Juan del Encina, 1495/1496]**
Compad**ré**, deb**és** sabér / qué la más buen**á** mujér / rábia siémpre pór j****.

b. **Sixteenth century ['Yo con vos', J. Román, 16th C.]**
Qué cuandó me viérdes / d**ó**quier**á** que fuérdes, / yó con vós.

c. **Seventeenth century ['Al villano se la dan', Anonymous 17th C.]**
Año buéno, Réy, tuvímos / pórque **sé**mbrandó⁀én el suélo / ...

d. **Eighteenth century ['Seguidilla dolorosa de una enamorada', from Una mesonera y un arriero, Luis Misón, 1757]**
A los mont**és** me salg**ó**, me salg**ó**,..

e. **Nineteenth century ['Himno nacional de Chile', R. Carnicer & E. Lillo, 19th C.]**
Que,⁀o la túmba serás de los líbrés / O⁀el asílo contr**á** la⁀opresión.

f. **Twentieth century ['Canción de primavera', C. Colomer, 1980]**
La prímavéra, la prímaverá llegó / La prímavéra, que prónto se márchitó.

ii. **Across space**

a. **Spain ['La panaderita', folksong]**
A lá⁀entráda del puéblo ý⁀a lá salidá, / Háy uná panadéra, panáderitá, panáderitá.

b. **United States ['La indita de Plácida Romero', New Mexican ballad]**
¿Cuandó te vólveré⁀a vér?

c. **Mexico ['La adelita', song from the Mexican Revolution]**
Si⁀Adelíta quisiéra ser mi⁀espósa / Si⁀Adelíta fuerá mi mujér

d. **Cuba ['La guajira', folksong]**
Me déspiertan lás tojósas, / Sálgo⁀al cámpo cón el sól.

e. **Dominican Republic ['Obsesión', melodanza]**
Si⁀al már escúcho / dicén las ólas / tu dúlce nómbre

f. **Puerto Rico ['La roca de la laguna', plena]**
Canário tiené⁀úna pléna / Que juéga cón la brisá del már.

g. **Guatemala ['Los carpinteros', action song for children]**
Déspues cóge sú martíllo / Ý las puértas vá⁀a clavár.

h. **El Salvador ['Las cinco vocales', children's school song]**
El paló de en médio és mas chíco cómo vés.

i. **Honduras ['Canción del marinero hondureño', folksong]**
Va dejándo⁀uná ladína,⁀una rúbia,⁀una moréna, y⁀ uná docéna quedán diciéndo.

j. **Nicaragua ['Credo', contemporary popular]**
Lás estrellás y la lúna / Lás casitás, las lagúnas

k. **Costa Rica ['¡Ay, Tituy!', folksong]**
Acordáte del centávo / Que te dí el añó pasádo.

l. **Panama ['Adiós, florecita blanca', folksong]**
¿Pára qué quierés caríño / Sí lo déspreciásté⁀ así...?

m. **Colombia ['Himno nacional de Colombia', national anthem]**
Lá libertád sublímé / Dérrama lás aurórás / Dé su⁀invencíble lúz.

n. **Venezuela ['Himno nacional de Venezuela', national anthem]**
¡Abájo cadénas! / Gritába⁀el señór, / Grítaba⁀él señór.

o. **Ecuador ['Pirusa', folksong]**
¿Qué te párece Pirusá / ... / Ló que nos ésta pásandó?

p. **Peru ['Llanto del indio', yaraví]**
Sómos los índios rójos puro⁀Atahuálpa / Que⁀én nuestra⁀atrícion sómos comó Sigálpa.

q. **Bolivia ['De blanca tierra', yaraví]**
De blanca tierrá hemos venído, / De nuestra tierrá hemos venído

r. **Paraguay ['La tejedora de Ñandutí', lace-making song]**
Aquí bien léjos de nuéstra pátria, / Tierrá del máte, Tierrá del máte y⁀ázahár.

s. **Uruguay ['Estilo', estilo]**
Résuena⁀el nómbre de Artígás, cómo⁀un ecó misterióşó.

t. **Argentina ['Bailando el gato', gato dance song]**
Cuandó los san juanínos / bájan el água

u. **Chile ['Ha nacido en un portal', Christmas carol]**
Ha nácido⁀en un portál / Llenitó de teláráñas / Entré la mula⁀y el buéy

v. **Judeo-Spanish ['Očo kandelikas', Yugoslavia, 20th century]**
Siéte kándelíkás / Očó kandélas pára mí.

iii. **Across genres**

a. **Pop rock ['Palabras, promesas']**
Palábras qué naciéron / Soló por tí / Pará sentír / Tus bésos.

b. **Folksongs ['Mercedita']**
Así nació / nuestró querér / con ilusión / con múcha fé / peró no sé

c. **Children's folksongs ['El patio de mi casa']**
El pátio dé mi cása / Es partículár / Cuandó llevé se mója / Cómo lós demás.

d. **Modern children's songs ['Muy bien, Tomás']**
Nó⁀encontró jardínes dónde póder jugár, ni⁀ éncontrába⁀amígos con quien póder hablár

e. **Villancicos ['Rin, rin']**
Mariá, Mariá / Ven acá volandó / Que los páñalillós / Los están royendó.

f. **Hymns ['Canten con alegria']**
Víno pará salvárlas peró sufriéndo cruentó dolőr.

g. **Popular songs ['Clavelitos']**
Mocíta, dáme⁀el clavél / Dame⁀el clável de tu bóca.

h. **Marches/National anthems ['Himno nacional de Venezuela']**
¡Abájo cadénas! / Gritába⁀el señór, / Grítaba⁀él señór.

i. **Classical Art Songs ['Jota', Manuel de Falla']**
Y,⁀áunque no quiéra tu mádre /Adiós, níña,⁀hastá mañána.

j. **Opera ['La vida breve'], Manuel de Falla]**
Pa désengáñós de⁀amórés / No⁀hay náda comó la muérté...

k. **Zarzuelas ['De España vengo', Pablo Luna y Carné, El niño judío]**
Que los téngo metíos dentró del álma / y que són los ojázos de mí gitáno.

l. **Gregorian chant ['Misa hispana']**
Tu que quitas el pecado del mundó / danos la paz.

m. **Ballet ['El amor brujo']**
Málhayá los ójos négros / que le alcánzarón a ve̋r

n. **Flamenco ['Chirigotas de Cádiz']**
Báta dé lunáres, batá de lunáres / Esá fue la⁀ heréncia que déjo mi máre

o. **Political songs ['El preso número nueve', Mexican Revolution]**
Yó se que⁀allá⁀en el ciélo / El ser súpremo nós juzgará.

To ensure that the stress-shifts in all of these song-samples are ones permitted uniquely in music, we have excluded from consideration all changes in accent which involve monosyllabic function-words that are stressable in citation form and so might be argued to bear stress at some stage of a phonological-phrasal derivation. Thus, for example, although stressing of the clitic **la** 'her' in a phrase such as that found in (3) is almost certainly deviant in any spoken dialect of Spanish, we have excluded such instances from our musical data because **la** does exist independently as a word and so can bear stress in its citation form.

(3) Lá tengo...
'I've got her...'

Moreover, in order to include only the most convincing evidence as part of our sample, we have also refrained from considering virtually innumerable cases of the types illustrated in (4) and (5).

(4) ...salidá⁀y... vs. ...salida⁀ý...

(5) ...Pánama... vs. ...Pànamá...

In (4), it could potentially be argued that the unusually-stressed final syllable of **salidá** is actually unstressed but receives its accent via coalescence with the following conjunction y - which could here itself be exceptionally stressed for the previous reason mentioned. That is, while stress would not normally be expected to fall on either of the two vowels in hiatus in **salida** y 'exit and', the fact that the latter word is monosyllabic has kept us from including this entire example - and all others like it - in our data set.

Cases like that in (5) have been excluded on other grounds. Although **Pánama** is not an acceptable pronunciation for oxytonic **Panamá** 'Panama' in normal Spanish speech, it is nevertheless generally recognized that in the spoken language a non-primary stress may fall two syllables away from main stress within a word. Based on this, we do not wish to treat musical accent-shift from primarily- to secondarily-stressed syllables on a par with accent-shift from stressed to completely-unstressed syllables. Such a decision is reinforced by the fact that a movement of accent from primarily- to secondarily-stressed syllables is also occasionally encountered in English song-texts, whereas such a movement from stressed to completely-unstressed syllables is - as we have already said - virtually unattested in English songs. Thus, for example, the English name **Clémentine** interchanges its primary and secondary stresses in sung **My dárling Clèmentíne**, but a reaccentuation like ***Cleméntine** is both bizarre and vanishingly rare.

Even with all the omissions just mentioned, though, the frequency of musically-conditioned stress-shift in Spanish is truly staggering - to the point where we are now surprised to find a song which lacks it entirely, rather than surprised to find a song which has it at all. Indeed, most texts are sung with more than one transfer of accent; many songs have multiple shifts in a single line, and some verses even include two reaccentuations within one word, like **péqueñá** instead of **pequéña**. Forms of this last type show that the phenomenon at issue here really is better characterized as deaccentuation plus reaccentuation than as stress-shift. That is, musically altered items like **péquená** are best analyzed as the result, not of somehow moving one original linguistic stress to two separate places in a word, but rather of deaccentuating an original linguistic stress that does not correspond to a musical accent and of newly accentuating two other syllables where stress is required by a song's own rhythm.

In order to cement these conclusions regarding how musical stress-shift occurs in Spanish to a degree totally unimaginable for English, we present in (6) the full Spanish text of the song 'Vientos del pueblo me llevan', originally written by Miguel Hernández-Aschero, as performed by the Coro Popular Jabalón on the Dial Discos tape **Cantos de la resistencia española**. Unusually stressed syllables which cannot appeal to the exceptions illustrated in (3-5) have been capitalized, and accentuation corresponding to regular Spanish orthography has not been disturbed.[2]

(6)
Vientos del pueBLO me
llevan,
vientos del pueBLO me
arrastran,
me esparCEN el corazón,
y me aventan la garganta.

Los bueYES doBLAN la
frente,
impotentemente mansa,
delanTE de los castigos:
los leones la levantan.

No soy de un pueBLO de
bueyes,
que soy de un pueBLO que
embargan
yaCImientos de leones,
desfiladeROS de garras
y cordilleRAS de toros
con el ORgullo en el asta.
Nunca MEdraRON los bueyes
en los páramos de España.

Quién habló de echar un
yugo
sobre el cuello de esta
raza?
Quién ha puesto al huracán
jamás ni yuGOS ni trabas,
ni quién al raYO detuvo
prisionero en una jaula?

AsTUrianos de braveza,
vasCOS de pieDRA blindada,
vaLENcianos de alegría
y castellaNOS de alma,
labrados coMO la tierra
y airoSOS coMO las alas;
andaluces de aceitunas,
naciDOS enTRE guitarras
y forjados en los yunques
torrenciales de las ansias;
extremeños de centeno,
galleGOS de lluvia y calma,
catalanes de firmeza,
araGOneSES de casta,
murcianos de dinamita
frutalmente propagada,
leoneses, navarros, dueños
del hambre, el sudor y el
hacha,
reYES de la minería,
SEñoRES de la labranza,
hombres que entre las
raíces,
como raíCES gallardas,
vais de la vida a la muerte,
vais de la nada a la nada:
yuGOS os quieREN poner
genTES de la hierba mala,
yuGOS que habéis de dejar
roTOS sobre sus espaldas.

Crepúsculo de los bueyes
EStá despuntando el alba.

Los bueYES mueREN vestidos
de humildad y olor a cuadra:
las áGUIlas, los leones
y los toros de arrogancia,
La agonía de los bueyes
tiene PEqueÑA la cara,
la del animal varón
toda la creación agranda.

Si me muero, que me muera
con la cabeZA muy alta.
Muerto y veinte veces
muerto,
la boca conTRA la grama,
tendré apretaDOS los dientes
y deCIdiDA la barba.

CantanDO espero a la muerte,
que hay ruiSEñoRES que cantan
enciMA de los fusiles
y en meDIO de las batallas.

In providing this admittedly rather extreme example of multiple stress-shift in a Spanish song-text, we wish not only to give concrete exemplification to that phenomenon but also to emphasize that - as we hope all those who are able to listen to the music will agree - neither the linguistic nor the musical integrity of the work in question suffers from the many unusual accentuations and deaccentuations which it shows vis-à-vis normal spoken Spanish.

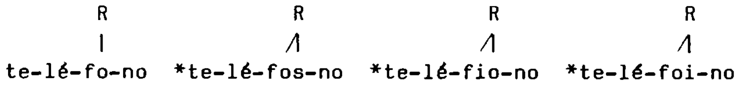
The stylistic - that is, phonetic - consequences of musically-conditioned stress-shift in Spanish, then, are quite substantial. But its linguistic - that is, phonological and morphosemantic - consequences are no less considerable. In numerous instances, important stress-marked distinctions are leveled in Spanish songs, so that there can be apparent shifts between, say, **levánto** vs. **levantó** 'raise (first-person singular present indicative vs. third-person singular preterite)' and **déje** vs. **dejé** 'leave [behind] (first- or third-person present subjunctive vs. first-person singular preterite)'. And other changes like those in (7) lead to antirhythmic stress-patterns which constitute egregious violations of the conditions on syllable-structure and stress proposed by Harris 1983 in his multilinear analysis of Spanish phonology and morphology.

(7) Me **déspiertan** las tojósas... ['La guajira', Cuban
'The pigeons awaken me...' folksong]

Yacímientos de leónes,... ['Vientos del pueblo me
'Fields of lions,...' llevan', song of the
Spanish Resistance]

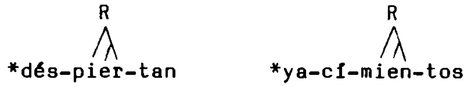
As the diagrams in (8) illustrate, antepenultimate stress is not possible in spoken Spanish words whose penult contains a branching rhyme.

(8)



But the diagrams in (9) show that, in Spanish song-texts, the shifted antepenultimate stress-pattern assigned to each of **déspiertan** and **yacímientos** creates not only a branching rhyme in their penults but in fact a rhyme with three branches.[3]

(9)



But why does - and how can - this stress-shift occur so frequently in Spanish, and why not in a language like English? One possibility - essentially the null hypothesis - is that whether or not a language allows musical stress-shift in its song-texts is simply a parameter of linguistic variation that must be arbitrarily stipulated. On this view, however, it is a complete accident that stress is shifted frequently in Spanish songs but rarely in English songs, rather than vice versa. Such an account is therefore surely to be rejected in favor of any alternative which explains the differential presence or absence of musical stress-shift in a language on the basis of structural features of that language. A first hypothesis which immediately comes to mind in this regard is that the notions of "syllable-timing" and "stress-timing" play a crucial role here. That is, if syllables are isochronous in Spanish, then shifting their stress should not affect the quantity and quality of the vowels in them to the extent that it would in English. Indeed, it might even be suspected that musical stress-shift should have no effect at all on the nature and duration of Spanish vowels. Furthermore, the inherent rhythmicity of most music might even encourage the imposition of such a minimally-interfering musical stress-pattern (and even musical stress-timing) on song-texts, since the timing of spoken-language stresses in Spanish is often claimed to be linguistically irrelevant.

This is surely a pertinent consideration, but it is completely undercut by a number of factors. To begin with, instrumental studies have failed to confirm the auditory impression of isochrony for Spanish syllables. For example, Dauer 1983 presents evidence that while, in Spanish, 'the syllable structure, lexical composition, processes of reduction [,] and phonetic realization of stress militate against the impression of regularity of interstress intervals', it in fact turns out that, physically, interstress intervals in Spanish are no less isochronus than in English. But even if this were not the case, the differential shift-ability of stress in Spanish vs. English songs still could not be attributed to the latter's being allegedly stress-timed.

This is because we have also gathered substantial evidence for musical stress-shift in Portuguese, which is now also usually said - like English - to be stress-timed (see, for example, Major 1985). Thus, as shown in (10), even totally-reduced vowels of Brazilian Portuguese may receive musical-rhythmic stress.

(10)	Não **falé** do médo	'Don't speak of the fear
	Qué **temós** da vída	we have of life
	Não pónha o dédo	Don't stick your finger
	Ná **nossá** ferída	In our wound (i.e. cut us to the quick)'

['Cartomante', Brazilian rock]

To mention just the first instance here, the orthographic **e** of **fale** 'speak!' receives a pronunciation with reduction to a high front vowel - [falí] - despite the shifted stress which it bears in the present musical context. Similar phenomena involving stress-shift onto reduced syllables obtain in Catalan, which is also usually said to differ from Spanish in being a stress-timed language.

Thus, as demonstrated by the partial song-text in (11), the Catalan reduced vowel [ə] can also bear musically shifted stress - as in the first and third lines of the Christmas carol given there.

(11) A vintícinc **dé** Desémbre, fúm, fúm, fúm,
Ha nascút un mínyonét ros í blanquét, ros í blanquét,
Fill de **lá** Vergé María, n'és nat én una éstablía, fúm, fúm, fúm!
A vintícinc **dé** Desémbre, fúm, fúm, fúm,
Es el día **dé** Nadál molt príncipál, molt príncipál,
Quan eixírem **dé** Matínes **fárem** bónes réstolínes, fúm, fúm, fúm!
['Baile de Nadal', Catalan Christmas carol]

In these cases, monosyllabic **de** 'of' and **la** 'the (feminine singular)' are sung with the stressed schwa that characterizes their citation-forms in the Barcelona dialect. On the other hand, in lines three and six of this carol, respectively, both **Vergé** 'Virgin' and **fárem** 'we will make/do' show both accentuation of a normally-unstressed vowel and deaccentuation of a normally-stressed and non-reduced full-valued vowel in favor of schwa. Finally, beyond Romance, it appears that musical stress-shift is also common in songs with texts in Russian - which is usually alleged, along with English, to be a prototypical stress-timed language, and certainly is one with widespread vowel-reduction conditioned by variable stress-placement.

Still, to bring in such additional facts is not to suggest that all or even any other Romance languages behave anything like Spanish in the relative frequency of stress-shift in their songs. Rather, we submit - as a second hypothesis - that the crucial consideration in the presence vs. absence of musical stress-shift in a language is the relative availability of certain word-types which facilitate or at least accommodate the matching of texts to tunes.

Thus, speakers of Lusitanian Portuguese, for example, have available to them a schwa which can be retained or omitted as needed to achieve a matching between linguistic and musical stress, as in a myriad of lexical items such as **el(e)** 'he/him', **príncip(e)** 'prince', **verd(e)** 'green', **esqueci-m(e)** 'I forgot', etc. Hence, while musical stress-shift does occur in that language, its frequency does not begin to approach the near-ubiquity of the phenomenon in Spanish. Similarly, stress is occasionally shifted for musical-rhythmic reasons in French song-texts, but this strategy is made less pressing because French speakers are likewise free to include or omit their schwa (or, more precisely, their [œ]) in such a way as to match texts to tunes without altering the stress-patterns of normal spoken French. And, of course, the usual final stress of French really has only a phrasal, rather than a lexical, domain, anyway. In the two excerpts in (12), it can be seen how in French a word like **comme** 'like' can be used to fill one syllable or two, depending on the needs and desires of a composer.

(12) a. Née au prin-temps, **com-me** la ro-se,...
'Born in the spring, like the rose,...'
['La mère Bontemps', Anonymous French]

b. Lui, **comme** nous, il est seule-ment / De
pas-sage...
'He, like us, is only / Passing through...'
['De Passage', Graeme Allwright]

Finally, of the Romance languages and dialects whose song-texts we have investigated more than cursorily, it seems that musical stress-shift also occasionally occurs in Italian, especially in older texts - such as those from the fourteenth century. However, even though Italian is generally said to be, like Spanish, a prototypical syllable-timed language, it is remarkable how rarely linguistic stress is shifted for musical purposes in the whole vast literature of internationally-known Italian opera and popular songs. The reason for this is apparently that Italian is able to offer its speakers a choice among various stylistic or dialectal alternants as a way to ensure text-to-tune matching without much need for stress-shift. As shown in (13), Verdi's alternation in **Rigoletto**, between **Monteróne** vs. **Monterón** '[a count's name]' and **uómo** vs. **uóm** 'man' illustrates how this strategy works.

(13) **Rigoletto, Act I, Scene 1:**

Borsa, Rigoletto, Marullo, Ceprano, Courtiers:	**Monterone** !
Monterone:	Sì, **Monteron**.

Rigoletto, Act I, Scene 2:

Sparafucile:	Un **uom** che...
Rigoletto:	...L'uomo son io che...

These reflections concerning Romance languages other than Spanish - and non-Romance languages other than English - are extremely suggestive, then, and so appear to support the hypothesis that musical stress-shift is conditioned by the exigencies of matching linguistic texts to musical tunes (especially rhythmically) more than by an alleged difference between syllable-timing vs. stress-timing in languages. Because we have delved much less deeply into the song-texts of these languages other than Spanish and English, though, we now leave behind those other tongues for a closer and more contrastive look at the latter two - although we hope to learn further information about musical stress-shift in other Romance and even non-Romance languages as our research continues.

More than any of the languages just mentioned, English is freest in its ability to achieve a relatively easy matching of texts to tunes without recourse to musical stress-shift. This is mainly due to the large number of monosyllables and of surface stress-patterns among its words, although certain syntactic characteristics of English (such as so-called 'preposition-stranding') are also relevant here. Spanish, however, is most limited in its text/tune-matching, since its mainly polysyllabic words have rather restricted stress patterns, lack an optional schwa, and are by and large not currently subject to any other rules of apocope or the like. It is these factors of usually polysyllabic word length and predominantly fixed non-final stress which, we contend, are most responsible for musical stress-shift in Spanish. This hypothesis also explains why stress is likewise frequently shifted in song-texts of Brazilian Portuguese and even certain non-Romance languages which otherwise seem to be phonologically quite different from Spanish. That is, these languages similarly have, relative to English, a predominance of polysyllables with largely fixed stress.

The claim that Spanish and English differ radically in the usual length of their words can be based on more than impressionistic feelings and anecdotes. As shown in (14), such a contention is completely borne out by broad-based statistical/empirical studies of the vocabularies of the two languages.

(14) SPANISH/ENGLISH

	Mono-syllabic	Bi-syllabic	Tri-syllabic	≥Tetra-syllabic
a. Absolute frequency:	14%/88%	66%/11%	18%/1%	2%/0%
b. Basic vocabulary:	22%/84%	64%/15%	12%/1%	2%/0%
c. Parallel texts:	48%/77%	32%/20%	12%/3%	8%/0%

(14) presents three comparisons of the Spanish and English lexicons for a different length of words in terms of syllables; in each case, the figures for Spanish are on the left of the slash, and the figures for English to the right. (14a) compares lexical types based on studies of the two

hundred most frequent words in Spanish and English (cf. Juilland and Chang-Rodriguez 1964 for Spanish, and Carroll, Davies, and Richman 1971 for English). Notice that the most common words of Spanish are overwhelmingly polysyllabic, while those of English are overwhelmingly monosyllabic. (When function-words - which are normally unstressed - are subtracted, the length differences between Spanish and English words are even more extreme, although we have not presented the numbers here.) In (14b), Spanish and English are compared in terms of word types within semantically-basic core vocabulary - the 200-word version of the so-called 'Swadesh list', which is generally assumed to be particularly resistant to replacement. Here, again, there is an extreme polarization between Spanish polysyllabicity vs. English monosyllabicity. Finally, in (14c), we present the results of a comparison between word types in a particular text - two versions of the same song: the English and Spanish translations of a hymn.[4] Yet again, we see a diametric opposition between the two languages in terms of word-length as measured in syllables, although function-words here swell the number of Spanish monosyllables by a not insubstantial amount.

It must be emphasized, though, that word length in syllables is important in conditioning musical stress-shift in Spanish song-texts only because Spanish stress is relatively fixed. If Spanish stress were lexically free, word-length would be irrelevant (at least as long as word length did not exceed verse length). Thus, as matters actually stand, Spanish words are overwhelmingly non-oxytonic, and must frequently undergo stress-shift in songs if they are to conform to the frequent musical pattern of final stress within rhythmic units.

The proposal just made is bolstered by a related, third hypothesis which concerns the linguistic role - the semantic load - of word-stress in Spanish vs. English. As already mentioned earlier, it is admittedly true that Spanish has a very large number of minimal pairs of the types illustrated - for various phonological shapes of verbs - in (15).

(15)

INFINITIVE	1st sg. pres.ind.	3rd sg. pret.	1st/3rd sg. pres.subj.	1st sg. pret.
BUSCAR:	búsco	buscó	búsque	busqué
DEJAR:	déjo	dejó	déje	dejé
HABLAR:	háblo	habló	háble	hablé
LEVANTAR:	levánto	levantó	levánte	levanté

These forms indicate the contrastive significance of stress in the language and outnumber by far the mere handful of English noun/verb pairs like those in (16).

(16)

NOUN	VERB
cómbat	combát
cónstruct	constrúct
pérvert	pervért

Still, stress in Spanish - unlike that in English - is completely predictable in all verbs and in all but a few exceptional nouns, adjectives, and adverbs. Harris (1983:83-84) states, for example, that 'there is no such thing as an unpredictably or irregularly stressed verb form in Spanish', and he goes on to say that

> 'although stress may appear on any one of the last three syllables of a [nonverb form], not all possibilities are equally likely ... Penultimate stress is unmarked in vowel-final words ... Final stress is unmarked in consonant-final words ... Plurals of [some words] have proparoxytonic stress, as expected ... Otherwise, proparoxytone stress in consonant-final words is severely restricted...'

Thus, in English, the position of stress must be memorized for a large number of words and hence would be much less recoverable if shifted in song-texts. But Spanish stress, being so predictable (especially given contextual clues as to lexical category and identity) is recoverable even when moved for musical purposes - so recoverable, in fact, that previous analysts have apparently failed to notice musically-conditioned stress-shift in the language or at the very least have failed to comment on it, thereby overlooking its significance as evidence precisely for the predictability of stress in Spanish. Not so paradoxically, then, the very predictability of stress in spoken Spanish leaves this stress so relatively bereft of semantic import that speakers are free to move it around arbitrarily in song-texts seemingly without any adverse effects on intelligibility or aesthetic evaluation.

By way of conclusion, then, we can rank the three main hypotheses considered here in terms of their potential for explaining musically-conditioned stress-shift in Spanish songs. Given the presence of frequent musical stress-shift in Brazilian Portuguese and other languages which are alleged to be stress-timed, the least important motivating factor for shifts of stress in Spanish songs seems to be the putative fact that the language is syllable-timed - or, alternatively, that it lacks processes of vowel deletion and reduction like those found in English. Much more important is the consideration that - as just mentioned above - Spanish spoken stress is relatively fixed. Most crucial, though, is the combined effect of this fixed, usually non-final stress with the fact that Spanish words are overwhelmingly polysyllabic. Together, these two factors virtually require speakers to shift spoken stresses in songs in order to gain any degree of freedom in matching a given linguistic text to one or more musical tunes whose rhythms feature a main stress which is other than penultimate or antepenultimate within a phrase.

In this paper, we have barely scratched the surface of the vast range of linguistic and musical phenomena involved in stress-shift among song-texts of Spanish and other languages both inside and outside the Romance family. The present investigation in fact represents just a portion of a larger ongoing study. As part of this continuing research we also plan, first, to conduct experiments concerning the salience and acceptability of musical stress-shift to speakers and, second, to investigate a spoken speech-style - the one closest to music - where quasi-musically-conditioned stress-shifts actually also occur: namely, dramatic recitation and oratory.[5] In any case, though, we hope to have shown here that Hispanic and general Romance linguistics has much to gain by considering musical settings of linguistic texts to be something other than deviations from or even interruptions of normal language. Instead, this intersection of language and music is a rich source of evidence bearing on such central grammatical claims as those made by Liberman 1978 and others regarding the matching of phonological strings to intonational tunes and by Harris 1983 and other regarding the multilinear assignment of stress. Thus, for those who have an open mind about investigation into all kinds of evidence bearing on that mysterious phenomenon of human language which intrigues and challenges us all, we suggest that the study of issues like musically-conditioned stress-shift should evoke from grammarians, not a reaction like 'That's frivolous and irrelevant', but instead a response like 'Ah, they're playing our song!'

NOTES

* The authors contributed equally to the research, writing, and presentation of this paper. Their names appear here in alphabetical order.

[1] By 'stress-shift', we will here mean any change in the prominence relations among the syllables in a word. Later in this paper we devote some brief remarks to how such shifts should be represented formally within a generative phonological framework. We in fact plan to make this representational issue a central focus of our further research into musically-conditioned stress-shift.

[2] Considerations of space prevent us from giving an English translation here, but glosses for all of the untranslated passages in this paper can be obtained by writing to either of the present authors.

[3] Actually, Harris applies his blanket condition against branching rhymes in proparoxytonic words only to non-verb forms, excluding verbs solely because of verb-plus-clitic combinations like **límpiese** 'clean yourself (formal)' and **hagámoslo** 'let's do it'. However, musically stress-shifted proparoxytonic **déspiertan** comprises only one morpho-syntactic domain (namely, a monomorphemic verb-root plus suffixes but without clitics) and so is certainly ill-formed. And the reaccentuated sung form **yacímientos** is, of course, a noun and thus totally without parallel in spoken Spanish.

[4] The hymn used for this count is 'O Sacred Head, Now Wounded'/'Oh rostro ensangrentado', original words by Paul Gerhardt (1656), based on a Medieval Latin poem. Both the English and the Spanish versions are modern translations. The tune is 'Passion Chorale', by Hans Leo Hassler (1601), harmonized by J.S. Bach (1729).

[5] We are indebted to Dwight Bolinger for suggesting the latter area of investigation.

REFERENCES

Carroll, J.B., P. Davies, and B. Richman. 1971. The American Heritage word frequency book. Boston: Houghton Mifflin.

Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. Journal of Phonetics 11.51-62.

De la Barra, E. 1887. Elementos de métrica castellana. Santiago de Chile: Imprenta Cervantes.

Harris, J.W. 1983. Syllable structure and stress in Spanish: a nonlinear analysis. (Linguistic Inquiry Monograph Series, 8.) Cambridge, Mass.: MIT Press.

Juilland, A. and E. Chang-Rodriguez. 1964. Frequency dictionary of Spanish words. The Hague: Mouton.

Liberman, M. 1978. The intonational system of English. New York: Garland.

Major, R.C. 1985. Stress and rhythm in Brazilian Portuguese. Language. 61.259-82.

Real Academia Española. 1973. Esbozo de una nueva gramática de la lengua española. Madrid: Espasa-Calpe.

John Earl Joseph

First-Person Singular Inversion in French

1. Subject clitic (SCL) inversion

Most French dialects have a rule of SCL inversion to signal interrogation:[1]

(1) Vous prenez le bus   Prenez-vous le bus?
You take the bus   Take you the bus?

Dialects differ in how limited SCL inversion is: relatively unlimited in the conservative standard, it is restricted to fewer verbs and contexts as one moves away from the standard socially or regionally. The unstable, non-'core' nature of SCL inversion in French is evidenced by the tendency of this syntactic process to morphologize, given that many of the dialects which appear to have done away with SCL inversion have in fact preserved it, at least vestigially, in the particles **est-ce que** and **-ti** which took over the interrogative function (see Brunot & Bruneau 1949: 528-32).

The chain of events in the weakening of SCL inversion appears to begin with a phonological problem: first-person singular **je** /žə/ is the only personal SCL to terminate in /ə/, the deletable vowel whose phonemic status has long been controversial (see Anderson 1982a, Morin 1978, Selkirk 1978). The initial consonant of a syllable having /ə/ as its nucleus will reattach to the preceding syllable and /ə/ will be deleted, so long as the preceding syllable is not made to violate French phonotactic strictures. When not deleted /ə/ is normally realized as [oe]. It is the only French vowel unable to take regular final-syllable stress.[2] This means that in phrase-final inversion from SCL-verb to verb-SCL, the stress will be on SCL in all persons except 1s (and with the impersonal SCL **ce**, which with few exceptions is restricted to cooccurrence with the verb **être**).

(2) Où es-tű vs. Où sui̋s-je /sɥižə/
Where are you Where am I

Où est-i̋l vs. Où e̋st-ce /ɛsə/
Where is he/it Where is it

*Où e̋s-tu *Où suis-je̋

*Où e̋st-il *Où est-ce̋

Each verb-SCL combination is stressed just as though it were a lexical word. The deletion of /ə/ yields phonotactically unproblematic syllables in the case of **suis-je** [sɥiž] and **vais-je** [vɛž] 'go I', but not so with verbs from the largest class, those with infinitives in **-er**, whose 1s present indicative and subjunctive forms also end in /ə/.^{3} They are left with a non-syllable-final /ə/ that becomes [ɛ] by a general phonological rule of French, formulated thus by Anderson (1982a: 545):

(3) $$\left\{\begin{matrix}/e/\\ \\ /ə/\end{matrix}\right\} \longrightarrow [\varepsilon]\ /\underline{\quad\quad}\ C \left\{\begin{matrix}\#\\ \\ C\end{matrix}\right\}$$

Rule (3) can be seen in operation in numerous verbs like **acheter** 'to buy' with a stem /ə/ that sometimes receives stress:

(4) acheter /ašɔte/ [ašte]
j'achète /žašətə/ [žašɛt] 'I buy'

The resyllabification of /t/ in **j'achète** feeds the stressed /ə/ into rule (3). Similarly with the je-inversion forms:

(5)
je mange /žəmãžə/ [žəmãz] mange-je /mãžəžə/ [mãžɛž]
I eat
j'achète /žašətə/ [zašɛt] achete-je /ašətəžə/ [aštɛž]

The inverted forms of (5) posed problems for French orthography, but after much debate the spelling was standardized as **mangé-je, acheté-je.**

The inability of **je** to take stress posed still greater problems for another group of verbs, whose primary distinguishing feature, according to standard French grammars (Grevisse 1969: 665) and to the reactions of my native-speaking consultants,[4] is their monosyllabic stem:

(6) *prends-je */prãžə/ *[prãž]
take I

*bois-je */bwažə/ *[bwaž]
drink I

Presumably, a monosyllabic verb-SCL combination does not allow enough phonological length for the realization of rising interrogative intonation. Normally the rise is distinguished on the basis of the higher pitch of the final syllable relative to those preceding; in ***prends-je** no syllable precedes. As a result, je-inversion sometimes yielded verb-SCL combinations that were homophonous with unrelated nouns or with the monosyllabic familiar imperatives of different verbs:

(7) *sers-je / serge [sɛrž] *dors-je / d'orge [dɔrž]
serve I / serge sleep I / of barley

*mens-je / mange [mãž] *romps-je / ronge [rɔ̃ž]
lie I / eat break I / rumination

Commentaries from the 17th century (Oudin, Menage, T. Corneille, cited by Harmer [1979: 52-54]) indicate that such homophonies were a source of conscious discomfort to speakers of the time.

Yet the phonological factor of monosyllabicity is not sufficient to distinguish the class of verbs for which je-inversion is blocked, since the monosyllabic verbs in (8a) do admit je-inversion. Those in (8b) admit it also, though apparently in a smaller number of dialects.[5]

(8) a. ai-je dois-je dis-je
have I must I say I

fais-je puis-je sais-je
do/make I can/may I know/can I

suis-je vais-je vois-je
am I go I see I

b. veux-je want I — viens-je come I

Furthermore, grammars of French specify (without elaborating) that only monosyllabic verbs of the **present indicative** are blocked from **je**-inversion. The only other tense containing monosyllabic forms is the preterite (**passé simple**), and the verbs which have them form a subset of the larger set of verbs with monosyllabic present tense forms (**je prends/je pris, je suis/je fus**; but **je veux/je voulus**).

2. A typology of SCL inversion constraints

The rules described in section 1 involve the minimum of constraints on **je**-inversion, exhibited by dialects which have not restricted SCL inversion more broadly. Such dialects represent a level of superconservative written usage that may rightly be pegged as artificial in its degree of aloofness from developments in the living spoken language. Most dialects currently in use have extended the constraints, giving the distribution shown in (9). **Aimé-je** represents all /ə/-final verbs; ***prends-je** represents monosyllabic present indicative verbs except those of (8); **suis-je** represents all other verbs. **Prends-tu** represents SCL inversion in all persons other than 1s. (* = ungrammatical; ## = ?* and/or very rarely produced; # = acceptable but seldom produced):

(9) a. aimé-je; suis-je; *prends-je, prends-tu; est-ce que je prends, #est-ce que tu prends; ##je prends

b. #aimé-je; suis-je; *prends-je, prends-tu; est-ce que je prends, #est-ce que tu prends; #je prends

c. ##aimé-je; #suis-je; *prends-je, prends-tu; est-ce que je prends, est-ce que tu prends; #je prends

d. ##aimé-je; ##suis-je; *prends-je, prends-tu; est-ce que je prends, est-ce que tu prends; #je prends

e. *aimé-je; ##suis-je; *prends-je, #prends-tu; est-ce que je prends, est-ce que tu prends; je prends

f. *aimé-je; ##suis-je; *prends-je, #prends-tu; #est-ce que je prends, #est-ce que tu prends; je prends

(9a) is the superconservative dialect type described in section 1. Type (9b) characterizes standard written French, together with the conservative spoken dialects that are so highly influenced by it as to be almost derivative of it (see Joseph 1987). In these dialects **je-inversion** has become archaic with /ə/-final verbs, and cannot be used without occasioning a high degree of stylistic markedness.

Over the centuries, the restrictions on 1s inversion have spread to classes of verbs not falling under the putative phonological blocks regarding monosyllabic and /ə/-final verbs. Standard (not superstandard) dialects are divided roughly between types (9c,d). In (9c), **je-inversion** is rare with any verb, and is marked for formality of discourse. In (9d), **je-inversion** is further restricted to just a small set of fixed phrases (**puis-je**, etc.).

I have not determined to what degree types (9c,d) occur independently of further SCL inversion restrictions. Perhaps in part because of the asymmetrical paradigm created by the restrictions on **je-inversion**, many spoken dialects have extended the limitations. In type (9e), **je-inversion** is restricted to a small set of fixed phrases, and SCL inversion in other persons is rare and marked for formality. In (9f), the most restrictive type of all, SCL inversion occurs only within a small set of fixed phrases.

While it is generally maintained that use of **est-ce que** played a key role in the decline of inversion, the distribution of this particle indicates a more complex interrelationship. The relatively few users of (9a) feel that **est-ce que** is 'heavy' and best avoided except as an alternative to unacceptable je-inversions. In dialect types (9b,e) **est-ce que** has moderately low frequency. In (9b) it is in variation with inversion; in (9e) with intonation alone, or zero signal, or both. In types (9c,d) **est-ce que** is used with the greatest frequency.

The reader may be puzzled by what seems to be in chart (9) an incongruous mixing of data from the traditionally distinct realms of competence (*) and performance (#). Like most cases of dialect variation, this one is not stable over time. Chart (9) displays a temporal dimension just as much as a social-stylistic one, with (9a) all but obsolete and (9f) continually on the rise. In the most plausible current model of language acquisition, an individual speaker's competence is the result of two separate input modules coming into contact: one, the speaker's innate linguistic competence, also known as Universal Grammar (UG); two, the performance of other speakers, to comprehend which the learner must extrapolate a grammar. This internal grammar, constructed on the basis of others' performance and guided by the learner's innate 'knowledge' of UG, becomes his or her competence.

Although we are far from knowing how to postulate whether a given linguistic mechanism is or is not a feature of UG, we are likely to exclude from consideration any mechanism which occurs in only a small fraction of the world's languages. Ultan (1978: 222), who surveyed the interrogative mechanisms of thirty-eight languages as part of the Stanford Project on Language Universals, found that 'YNQ [Yes-No Question]-inversion appears to be a rather uncommon interrogative device, occurring in only seven languages (six of which are modern European) out of 38.' The seven were English, Finnish, French, Hungarian, Malay, Rumanian, and Russian. The distribution of languages in Ultan's sample does not reflect linguistic distribution worldwide (as does, by contrast, the sample used by Bybee [1985]); if it did, we would probably see even less evidence for postulating inversion as an inherent facet of UG. The alternative is that it derives in part or in whole from transmission between generations - i.e., by rule extrapolation from performance.

Thus in chart (9) we can understand the history of a form like **aimé-je** as follows: in the span of a 'generation' between dialect types (9a) and (9b) the means of posing a first-person singular question for an /ɔ/-final verb ceased to be a free variation between inversion and **est-ce que**. Children of generation (b) heard both question forms in their parents' output, and even though **aimé-je** may in some cases have been more frequent than **est-ce que j'aime**, the particular grammar of French which a majority of children in generation (b) extrapolated handled the **est-ce que** structure more readily, though none of these speakers would deny that

aimé-je was acceptable. Why their grammar handled **est-ce que** more readily is not yet knowable; whatever the reason, in the input to (b) speakers' competence, and subsequently in their own performance, the grammatical **aimé-je** was comparatively infrequent. Their performance became the input for generation (c). Having no access to (b) speakers' competence except through their performance, (c) speakers could not know that **aimé-je** was still fully acceptable in dialect (b), but naturally inferred from its distribution that it was marked. Hence it became still less frequent in dialect (c), so much so that generation (d), extrapolating a grammar from generation (c)'s output, did not internalize **aimé-je** at all.

There is no reason to doubt that the great majority of linguistic changes proceed in exactly this fashion, and that data of both competence and performance are necessary to the proper understanding of language diachrony and synchronic variation.

3. Are the je-inversion constraints phonological?

My comments on the data in section 1 are based on the traditional view that je-inversion is blocked for phonological reasons. The continuing prevalence of this view explains how it is that the intricacies of 1s are summarily ignored in the extensive literature on SCL inversion (e.g., Kayne 1972, 1975, 1983: 211ff.; Bailard 1981; Muller 1984; Safir 1985: 204-15). Yet no purely phonological treatment is able to account for the data, as the following pieces of evidence show:

a) Since the forms of (8) are acceptable while those of (6,7) are not (***bois-je** vs. **dois-je**), semantic or syntactic factors, or both, must be involved.

b) That monosyllabic verbs of the preterite tense can undergo je-inversion, while those of the present tense cannot, demonstrates that a morphological element is crucially involved.

c) The rules invoked are not supported by the overall phonology of French. Specifically, we have no parallel instance of word length determining any other syntactic behavior. Likewise,

d) even the primum mobile of the constraint, the unstressability of /ə/, is leaky, given examples like **prends-le** (see note 2).

e) In every dialect type of (9) except perhaps (9d), the distribution of **ce** is much more highly restricted than is that of the phonologically more marked **je** (/s/ has one phonological mark in the Chomsky & Halle [1968] schema, while /ž/ has four), an indication that this block is not primarily phonological.

f) Recent work in the GPSG framework has refocused attention on the Principle of Phonology-Free Syntax (PPFS), originally formulated by Zwicky (1969). Since models of generative grammar have always held that the output of the syntactic component is the material upon which systematic phonetic rules operate, the PPFS forbids analyses in which phonological considerations motivate or block syntactic operations. Pullum & Zwicky (1984) argue that the PPFS should be maintained in its strongest form, and provide a case study of the distribution of English inflectional (**-er/-est**) and periphrastic (**more/most**) comparatives and superlatives. It has long been maintained (e.g. by Sweet [1891: 326-7] and more recently by Huddleston [1973: 353]), that this distribution is a function of the phonological length of adjectives, and as such would constitute a PPFS violation. Pullum & Zwicky show that the putative phonological generalizations regarding this distribution admit so many exceptions as to lack any explanatory power; they conclude that

> the division of adjectives into inflecting and periphrastic subcategories turns out to be a matter of arbitrary lexical conditioning. The tendency for one subcategory to contain shorter stems than the other is explicable historically and is not grammatically relevant. We have found that this sort of situation is typical of the various putative phonologically constrained LRR's [lexical redundancy rules] that have been suggested for English or other languages. (114)

Although the PPFS needs more demonstrations in more languages before acceding to the status of a law, it should nevertheless be a general property of grammar construction that any account which predicates syntactic rules on phonological factors is ipso facto more 'expensive' than an alternative account of comparable (or even slightly greater) complexity that does not invoke such a crossing of levels.[6]

In the case of first-person singular inversion in French, an alternative account would have two primary phenomena to explain: the failure of certain verbs to undergo the process in any dialect type (***prends-je**), and the rise of a block against the process in /ə/-final verbs (***aimé-je**). Because the latter suggests a possible approach to the whole problem, I shall consider it first.

4. Je-inversion in /ə/-final verbs

In section 1 I mentioned that 17th-century commentators expressed discomfort with the homophony of je-inversion forms of certain verbs with other words, as shown in (7). Another of their principle concerns was with the 'illogical' je-inversion that occurred for some of these same verbs:

(10) servé-je (for *sers-je)   dormé-je (for *dors-je)
mənté-je (for *mens-je)   rompé-je (for *romps-je)

The commentators considered these forms solecisms, created by analogy with 'proper' forms like **aimé-je**, to replace the 'logical' but 'poor sounding' forms in parentheses (Oudin, Vaugelas, Menage - and later Littré and Brunot - all cited by Harmer [1979: 52-54]). None of the commentaries I have read makes note of the fact that, not only are the forms of (10) identical to the present subjunctive of each verb, they are also homophonous with the imperfect tense form.

(11) [sɛrvɛž]   [dormɛž]   [mãtɛž]   [rɔ̃pɛž]
servais-je   dormais-je   mentais-je   rompais-je
servé-je   dormé-je   menté-je   rompé-je

Nor do the grammars mention that these same homophonies - plus a third, with the preterite - characterize the je-inversion forms of the **-er** verb class.

(12) [emɛž] aimé-je (indic. & subj.)/aimais-je/aimai-je
[aštɛž] acheté-je (indic. & subj.)/achetais-je/
achetai-je

I can think of no explanation of the forms in (10) more plausible than this: the speakers who generated them frequently heard je-inversion forms in the **-er** verb class, and in order to process them constructed in their internal grammar a rule of approximately the following shape:

(13) In 1s-inversion, use the corresponding subjunctive/
imperfect.

Rule (13), which has at least as much explanatory simplicity as a derivation from within the present indicative paradigm, also has surface evidence in the fact that the phonetic forms of (10) previously existed and coexisted in the speaker's experience as imperfects and subjunctives. We do not have enough evidence to know just how far such speakers constrained the rule, nor how they handled forms like **fussé-je** (see note (3)); but then no single, simple rule can handle both **fussé-je** and **menté-je**.

The later date of attestation of the forms in (10) is not surprising, since the rule reinterpretation required a lapse of time; their ephemerality is perhaps more significant. The forms began to be used precisely at the historical moment when the forces of linguistic control in France had reached a pinnacle of power and linguistic consciousness (see Joseph 1983). Not only did the forms never occur in many dialects, but speakers who did use them met with disdain and ridicule. Their use was thus very highly restricted, and if they have not died out entirely, just a handful of attestations can be found in the twentieth century.

What this indicates is that rule (13) was not generally perceived as being operative. Nevertheless, it was the case that homophony of the present, imperfect, and preterite indicative, and the present subjunctive, characterized the je-inversion forms of the largest and only productive verb class in dialect type (9a). And rule (3) created an additional homophony between the future and conditional je-inversion forms of every verb in the language.

These are not the sort of homophonies that can be easily tolerated. First-person interrogation is, a priori, a marked domain of discourse, given that the function of interrogation does not lend itself to the first person in a natural way, only in a derived way. Although in morphology marked domains are characterized by the neutralization of formal distinctions (see Joseph 1986), marked discourse domains tend to be those that receive maximal foregrouding: on the rare occasions when one asks a question about oneself, it is especially crucial whether one **was** or **is** whatever is being wondered about.

In sum, I am postulating that homophony of the present, imperfect, and preterite tenses - a morphological/semantic consideration, rather than a phonological one - motivated the decline of je-inversion in /ə/-final verbs. Furthermore, the fact that the phonetic form [emɛž] continues to be generated as the imperfect **aimais-je** long after it has ceased to be generated as the present **aimé-je**, is yet another argument to add to the list in section 3 against a phonological block.

5. The type *prends-je

Having turned up a plausible morphological explanation for one of the constraints on je-inversion, we shall consider whether an explanation at the same level can be devised for those non-/ə/-final verbs for which je-inversion is unacceptable even in dialect type (9a). Is there a way of circumscribing them morphologically, syntactially, or semantically that is superior to the traditional analysis, which ascribes their syntactic behavior to their phonological length and predicates a list of exceptions that includes the most textually frequent verbs of the lot? This traditional form of the oldest and most widespread restriction on SCL inversion, the only one operative in (9a), may be expressed as follows:

> (14) je-inversion is inoperative with present-tense monosyllabic verbs, other than a small number of exceptional ones.

Although Anderson's query 'Where's morphology?' (1982b) awaits a satisfactory answer (as Scalise 1984 demonstrates), it is clear that morphological operations precede the phonological component. Morphological parameters for je-inversion would involve at most lexicon and syntax, and so would be 'cost-effective'. If we consider (as is normal) present indicative to be [U tense/mood], then we can say that those verbs which do not undergo je-inversion also do not undergo any derivational or inflectional morphological processes. Consider the following examples:

(15) *prends-je pris-je prendrai-je comprends-je finis-je

***Prends-je** is unmarked for tense or derivation. **Pris-je** is marked for past tense, **prendrai-je** for future tense, both are inflectionally derived and can undergo je-inversion. **Comprends-je**, **apprends-je**, **surprends-je**, and every other affixational derivation can undergo je-inversion; as can the large group of verbs like **finir** containing an original inchoative affix -is(s-) (cf. **nous finissons**, beside **nous prenons**), which though no longer productive still marks verbs as lexically derived. Outside of the productive **-er** class, the French lexicon has almost no polysyllabic verbs that are not formed derivationally from monosyllabic ones.[7]

Since /ə/-final verbs in inversion undergo a change that is in part morphological (***aime-je** ⟶ **aimé-je**), we can posit as the first approximation of an alternative to (14):

(16) je-inversion is unblocked by any morphological process of derivation or inflection.

This rule gives the correct results for all but the 'exceptional' verbs of (14), which are listed in (8) and repeated for convenience as (17):

(17)
ai-je dois-je dis-je fais-je puis-je sais-je
suis-je vais-je vois-je veux-je viens-je

Morphologically, it is true that these verbs belong to some of the most irregular and suppletive paradigms of the language; yet **prendre** is quite as irregular as **dire** or **faire**, and **boire** (***bois-je**) follows closely the pattern of **devoir** (**dois-je**).

On the syntactic level, however, we note that (17) contains virtually all of the verbs able to function as tense auxiliaries (avoir, être, aller, venir), modal auxiliaries (devoir, pouvoir, savoir, vouloir), or causative (faire). Any grammar of French has to designate these verbs specially for syntactic patterns having nothing to do with ls inversion, so that to mark them also for the latter involves very low cost. Interestingly, these are precisely the categories of verbs to which inversion is restricted in English (the originally causative **do** having lost that function and become a sort of dummy auxiliary for inversion; see Lightfoot 1979: 115-120; Battistella 1985). This cross-language parallel strengthens the grounds for asserting that the verbs of (17) are bound by syntactic features.

The verbs **dire** and **voir** belong to semantic categories with inherent syntactic peculiarities not only in French, but in many of the world's languages. Munro (1982), in her study of narrative verbs, speaks of their 'nontransitive' and often auxiliary-like behavior. **Say** and other narrative verbs are marked by the property of subject-verb inversion in both French and English: **dit le président, said the president**. As for perception verbs like **voir**, they resemble the modal auxiliaries in their ability to be followed directly by a dependent infinitive: **regarder, écouter**, and **entendre** all share this syntactic property with a few other verbs, and it may be a clue to precisely what syntactic feature conditions je-inversion in verbs undergoing no morphological process.[8] **Voir** gets intrinsic low marks on at least three parameters of the Hopper-Thompson (1980) transitivity scale, for volitionality, kinesis, and affectedness of O[bject]; non-intrinsic clausal context can only drive its transitivity lower still. Its historical relationship with syntactically complex verbs like **seem** and **appear** is also well known.

Even if, under further scrutiny, one or more of the syntactic parameters I have postulated for je-inversion should fall, the case would still be stronger than that made by Pullum & Zwicky against the phonological determination of English comparatives and superlatives, which does not deny the existence of broad-based phonological correlates, but only shows that they cannot be the primary determinant. For now, I shall reformulate rule (16) as (19a):

(19a) je-inversion, blocked in full verbs (those not categorized with the syntactic properties of auxiliaries), is unblocked by any derivational or inflectional morphological process.

To paraphrase my earlier citation from Pullum & Zwicky: the division of verbs into [+ je-inversion] and [- je-inversion] subcategories turns out to be a matter of lexical conditioning and syntactic processing.

6. A history of je-inversion constraints

Rule (19a) is operative for dialect type (9a). Its subsequent history is one of increasing strength, as revealed by the progressive appearance of dialect types (9b-f). The operative rule at each of these levels is approximately:

(19b) In the output of (19a), present-imperfect tense homophony is to be avoided.

(19c) Constraint (19b) is strengthened and begins to be reinterpreted as a general block against **je**-inversion.

(19d) je-inversion is essentially blocked.

(19e-f) SCL inversion progresses toward blockage.

If the full elimination of SCL inversion comes to pass, it will signal the completion of a still larger historical development, the loss of subject-verb inversion. Both Old French and Old English signaled interrogation by inversion of the subject, with nouns as well as pronouns, as Italian and Spanish continue to do. In perspective, it seems that both French and English have been gradually restricting and eliminating this syntactic device for roughly a thousand years. French began by restricting which subjects could undergo inversion, English by which verbs. Just as the restrictions have grown in French, the grammar of English continues to eliminate the [+inversion] feature from verbs. Within this century the verb **have** has, in many dialects, lost [+inversion] in its use as a full verb, though retaining it when used as an auxiliary: **Have you seen the paper** is still normal, but **#Have you the paper** is increasingly less so. Even the stability of the former is suspect, given the widespread tendency to **drop** the auxiliary in such sentences (**You seen the paper?**). Auxiliary dropping further weakens the statistical frequency of subject-verb inversions in the input to children's language acquisition, and can only aid in the long development toward a solidly fixed SV order.

In French, the growing loss of SCL inversion may possibly be connected with another long-term trend, the move from suffixational to prefixational morphology (DeCesaris 1986; Fleischman 1982: 127). If so, the consequences for the language's government and binding configurations are certain to be enormous; but they will have to be saved for a subsequent study.

NOTES

[1] SCL inversion also occurs in a very few non-interrogative functions, such as in exclamations and after certain sentence-initial adverbs (like **peut-être**). The facts of variation and distribution in these peripheral cases do not vary from those of interrogation.

[2] /ə/ can receive stress if it is the vowel of a monosyllabic word in absolute final position: **prends-le** 'take it', **sur ce** 'on this', **parce que** 'because' (Anderson 1982a: 539). SCLs are not treated as independent words, but as the final syllable of the host word; thus **prends-le** but ***prends-jë**, and **sur ce** but ***est-cë**. / / is realized as [oe] in such cases (with some variation only in the case of **le**), even though /œ/ is raised from [œ] to [ø] in absolute final position.

[3] In addition to the present indicative 1s of **-er** verbs, the imperfect subjunctive 1s paradigm forms of all verbs ended in /ə/. They were processed exactly like the present tense forms shown in (5): **je fusse** 'I were' ⟶ **fussé-je** 'were I', etc. However, the imperfect subjunctive was already in its decline when the first constraints against je-inversion came into play, and its frequency in spoken usage was subsequently so low that we cannot consider that it played any measurable part in the development of the je-inversion constraints discussed below.

[4] I gathered very complete data from fourteen speakers, and partial data from nine others, during a stay in Paris and Normandy in December 1985. They represented a spectrum from working to upper-middle class, and with the exception of one, who had lived eighteen years in Béziers before moving to Paris, were natives of the areas in which I interviewed them. I found no significant regional differences in je-inversion that could not be attributed more cogently to social and educational level.

[5] Rosenberg et al. (1983: 248) state categorically that 'with the exception of nine forms [those of (8a)], inversion is generally avoided in the first person singular present indicative.' Grevisse (1969: 665) says that 'Toutefois l'usage admet l'inversion de **je** après quelques verbes très usités', then lists the verbs of (8a) plus **veux-je**, and in his examples includes a **viens-je**. My informants likewise showed less accord regarding the verbs of (8b).

[6] During the discussion period which followed this paper at LSRL, Richard D. Janda suggested that one might obviate the PPFS by positing that what is blocked in cases like ***prends-je** is not inversion, but rather cliticization. Since in French cliticization consists of two components, a syntactic one of placement and a morphophonological one of liaison/enchaînement, we could say that the latter is blocked, continue to posit a phonological motivation for the block, and not violate the PPFS. But French does not have any sort of general filter against clitics with liaison/ enchaînement blocked - I am thinking specifically of those whose hosts begin with '**h** aspiré' - so that to eliminate an inverted but not morphophonologically bound **je** would require a special phonological filter that would still have to apply after the syntactic cycle, thus taking us back to square one.

[7] **Ecrire**, an outstanding exception, might well be marked as a 'narrative' verb along with **dire** (see below).

[8] However, two other 'perception' verbs, also followable directly by an infinitive, are not able to undergo **je**-inversion: ***crois-je** 'believe I', ***sens-je** 'feel/ smell/hear I'.

REFERENCES

Anderson, S.R. 1982a. The analysis of French schwa. Language 58.534-571.

------. 1982b. Where's morphology? Linguistic Inquiry 13.571-612.

Bailard, J. 1981. A functional approach to subject inversion. Studies in Language 5.1-29.

Battistella, E. 1985. Markedness isomorphism as a goal of language change: The spread of periphrastic **do** in English. Lingua 65.307-322.

Brunot, F. and C. Bruneau. 1949. Précis de grammaire historique de la langue française. 3rd. ed. Paris: Masson.

Bybee, J.L. 1985. Morhphology: A study of the relation between meaning and form. Amsterdam and Philadelphia: John Benjamins.

Chomsky, N. and M. Halle. 1968. The sound pattern of English. New York: Harper and Row.

DeCesaris, J.A. 1986. A lexical analysis of the Catalan periphrastic past. Paper presented at LSRL XVI.

Fleischman, S. 1982. The future in thought and language. Cambridge: University Press.

Grevisse, M. 1969. Le bon usage: Grammaire française avec des remarques sur la langue française d'aujourd'hui. 9th ed. Gembloux: Duculot.

Harmer, L.C. 1979. Uncertainties in French grammar. Ed. P. Rickard and T.G.S. Combe. Cambridge: University Press.

Hopper, P.J. and S.A. Thompson. 1980. Transitivity in grammar and discourse. Language 56.251-299.

Huddleston, R.D. 1973. Rev. of Introduction to stratificational grammar, by D.G. Lockwood. Journal of Linguistics 9.349-356.

Joseph, J.E. 1983. Interferences of standardization on the transition from Middle to Modern French. Journal of Historical Linguistics and Philology 1.1.34-63.

-----. 1986. Isomorphism and value correlations: Evidence from the Latin and French verb phrase. ESCOL 1985: Proceedings of the Second Eastern States Conference on Linguistics, ed. S. Choi, D. Devitt, W. Janis, T. McCoy, and Z. Zhang, 120-131. Columbus: Ohio State University Linguistics Department.

-----. 1987. Eloquence and power: The rise of language standards and standard languages. London: Frances Pinter.

Kayne, R.S. 1972. Subject inversion in French interrogatives. Generative studies in Romance languages, ed. J. Casagrande and B. Saciuk, 70-126. Rowley MA: Newbury House.

-----. 1975. French syntax: The transformational cycle. Cambridge MA: MIT Press.

-----. 1983. Connectedness and binary branching. Dordrecht: Foris.

Lightfoot, D.W. 1979. Principles of diachronic syntax. Cambridge: University Press.

Morin, Y.-C. 1978. The status of mute 'e'. Studies in French Linguistics 1:2.79-140.

Muller, C. 1984. L'inversion du sujet clitique en français et la syntaxe du sujet. Lingvisticae Investigationes 8.335-362.

Munro, P. 1982. On the transitivity of 'say' verbs. Syntax and Semantics 15: Studies in transitivity, ed. P.J. Hopper and S.A. Thompson, 301-318. New York: Academic Press.

Pullum, G.K. and A.M. Zwicky. 1984. The syntax-phonology boundary and current syntactic theories. OSUWPL 29.106-116.

Rosenberg, S.N., M.T. Houston, R.A. Carr, J.K. Hyde, M.D. Moody, and E.W. Najam. 1983. Harper's grammar of French. New York: Harper and Row.

Safir, K.J. 1985. Syntactic chains. Cambridge: University Press.

Scalise, S. 1984. Generative morphology. Dordrecht: Foris.

Selkirk, E. 1978. The French foot: On the status of 'mute e'. Studies in French Linguistics 1:2.141-150.

Sweet, H. 1891. A new English grammar, logical and historical. Part I. Oxford: University Press.

Ultan, R. 1978. Some general characteristics of interrogative systems. Universals of human language, 4: Syntax, ed. J.H. Greenberg, C.A. Ferguson, and E.A. Moravcsik, 211-248. Stanford: University Press.

Zwicky, A.M. 1969. Phonological constraints in syntactic descriptions. Papers in Linguistics 1.3.411-463.

Marie-Odile Junker

Transitive, Intransitive and Reflexive Uses of Adjectival Verbs in French

This study on de-adjectival verbs will try to relate their morphological subclasses to their syntax and semantics. We will see that diachronic facts shed light on the verb's syntactic primitive (transitive or intransitive); moreover, affix distribution and semantic restrictions correlate with the syntactic primitive; finally an account is provided for intransitive vs reflexive use of the verbs.

A. Two uses and two morphological subclasses for de-adjectival verbs

Verbs derived from adjectives, like **grandir** or **embellir** very often have two uses, an inchoative intransitive one (a) and a causative transitive one (b):

(1) a. il grandit
    he is growing

b. il le grandit
    he makes him/it taller

(2) a. il embellit
    he grows prettier

b. il l'embellit
    he embellishes him/it

An inspection of the list of all the verbs derived from adjectives reveals the existence of two morphologically different subclasses: those derived with the addition of a derivational affix (3) and those derived without such an affix (4):

(3) **embellir** and not ***bellir** (to embellish)

(4) **grossir** and not ***engrossir** or ***agrossir** (to grow bigger)

We will explore the consequences of those two modes of formation on the syntax and the semantics of the resulting verbs.

B. Initial hypothesis

The initial hypothesis we will discuss is:

(5) a. The presence of a derivational affix on de-adjectival verbs indicates that the verbs were initially (historically) transitive, with a causative meaning.

(5) b. Verbs derived from adjectives without such an affix are initially intransitive, with an inchoative meaning.

We suggest that a verb like **embellir** was initially derived with a causative meaning, i.e. it was transitive; historically, (2b) was the first syntactic use of the verb. The affix **em-** is a morphological trace of this initial syntactic use. On the other hand, the initial syntactic use of **grandir** is, according to (5b), the inchoative one: (1a).

In order to test the hypothesis in (5), one must investigate the historical evidence. For example, **embellir**, in its first use, is transitive: it is first attested in the XIIth century (the dates are 1100 and 1130 according to GLLF and TLF) with a causative meaning:

(6) a. 1100: il l'embellit

The first attestation of the intransitive use is not until 1671:

(6) b. 1671: il embellit

Grandir appears in French around 1260 with an intransitive, inchoative use: its first use is (7a):

(7) a. 1260: il grandit

The transitive use (7b), however, is said to be rare before the XIXth century and the earliest attestation we have for this use is 1460.

(7) b. 1460: il le grandit

C. The data

C.1. Data collection

450 deadjectival verbs were checked in all major dictionaries (Grand Larousse de la Langue Française: GLLF, Trésor de la Langue Française: TLF, Littré, Lexis, Robert, Godefroy, Huguet, Tobler-Lommatsch, etc.). The only dictionaries actually giving dates for different uses were the GLLF and TLF. However, the other dictionaries were useful to check meanings and most common uses.

In using the dictionaries we kept in mind the following:
1. the order in which the two uses are given may be arbitrary and therefore cannot be used as evidence.
2. though TLF and GLLF are the only dictionaries which give dates for each use, they do it in some sporadic ways: TLF excellent etymological and historical descriptions, but only only in its first volume (letter A/Af...). GLLF has good articles but only starting with letter C: most of the time, the dictionary does not state the dates of each syntactic use (transitive, intransitive) but restricts itself to a single date of the verbal form (a date which may even show an exceptional use).
3. if a syntactic use is attested, it may not be the one you are looking for: 'intransitive' uses may simply be absolute transitives (**la lampe éclaire...(la pièce)...** 'the lamp lights...(the room)...) and anything attested by a past principle should not be reliable (in Old French the prefix **en-** could be added to form the past participle of a **a-** or an unaffixed verb; for example, rougir ⟶ past participle: enrougi. i.e. there is no verb such 'enrougir').

The data are summarized in (8).

(8) Verbs: uses and affixes

affix	total	trans. only	intr. only	both	date: trans intr	date: intr trans	date n.a.	total: trans	total: intr
IFI	32	32	1	0				31	1
IS	145	145	0	0				145	0
A	72	53	0	19	16	0	3	72	19
DE	16	16	0	0				16	0
E	18	15	0	3	1	0	2	18	3
EN	23	17	0	6	5	0	1	23	6
0	104	58	12	34	1	19	14	92	46
other	40								

C.2. Comments

Affixes having an effect on verbal valency are:
1. the prefixes **a-, en-/em-, é-** and **dé-**; for example: enlaidir, élargir, dégourdir;
2. the affixes **-is-** and -(i)fi- preceding the inflectional affix and referred to as the suffixes **-iser** and **(i)fier**.
Ø symbolizes the lack of such an affix; the verb just has inflectional ending.

The affixes having no effect on the verbal valency and therefore excluded from this study are in the columm **'other'**. They are:

-ail-: grisailler (to turn grey)
-oy- : rougeoyer (to turn red)
-on- : grisonner (to turn grey)
dis- : discontinuer (to discontinue)
pro- : prolonger (to prolong)
sur- : surenchérir (to bid higher)
ré-, r-, re-: réhumidifier (to rehumidify)
and numerals: tripler, quadrupler, centupler... (to triple, to quadruple, to centuple)

In this study, 410 (out of 450) verbs were finally considered.

Total: shows the number of deadjectival verbs per affix. This total number of verbs is divided between those having the transitive use only (**trans only**), the intransitive use only (**intr only**) and those having both uses (**both**). The historical evidence is given in the columns called **date:** there are three possibilities: the transitive use can be attested earlier than the intransitive one (**date: trans intr**); the intransitive can be the earliest (**date: intr trans**) or the dates are not available (**date: not available**). The final two columns of subtotals show: **total of trans, total of intr.**

C.3. Results

The data reveals that: 1) only verbs without the affixes under discussion can be uniquely intransitive (see: intransitive only), and 2) if a verb has an affix it must minimally have a transitive use. There is only one verb which violates the latter generalization; this verb (**gâtifier**, to become senile) contains an affix but is uniquely intransitive.

The historical data - when available - fulfills 5a: all affixed verbs with both uses have their transitive use attested earlier than the intransitive one. Out of 19 **a**-verbs having both uses: 16 have their transitive use attested earlier than their intransitive one. No dates are available for the remaining 3 verbs; of the 3 **é**-verbs which have both uses, 1 is attested as predicted, 2 have non-available dates; for 6 **en**-verbs, 5 are attested earlier in the transitive use as well, 1 has non-available dates. There are no counter-examples.

Regarding (5b), things seem more complicated: the bulk of the historical data supports (5b) with respect to verbs having both uses; 19 of these verbs have their intransitive use attested first, there is only one counter-example (roussir).

Nevertheless 58/104 of the unaffixed verbs are uniquely transitive: we cannot maintain the claim that a verb without an affix necessarily has an intransitive syntactic primitive, as hypothesized in (5b). We must reformulate (5b) in light of the distribution of affixed and unaffixed verbs.

D. Distribution of affixed and unaffixed verbs

Some deadjectival verbs show complementary distribution with respect to affixes, for example (9):

(9)	a. faiblir	b. affaiblir
	to weaken	to weaken
	(intransitive)	(transitive)

Faiblir has only an intransitive use, whereas **affaiblir** has only a transitive one. Notice that both verbs have been cited since 1100 and that the presence of the affix is related to a specific use: the causative one.

Similar examples can be seen in (10), where the affixed verb never has an intransitive use. The intransitive use is expressed by the form without affix; this unaffixed verb can also be transitive, but this use is less common:

(10) Examples:

Unaffixed	Affixed	Adjective
Baisser	Abaisser	Bas
Chauffer	Echauffer	Chaud
Durcir	Endurcir	Dur
Forcir	Fortifier	Fort
Gauchir	Dégauchir	Gauche
Grossir	Engrosser/Degrossir	Gros
Hausser	Exhausser	Haut
Maigrir	Amaigrir	Maigre
Mincir	Amincir	Mince
Mollir	Amollir/Mollifier	Molle
Sècher	Assècher	Sèche
Tiéder	Attiéder	Tiède

In (10) affixed verbs are definitely marked with respect to transitivity whereas unaffixed verbs are unmarked. We can now reformulate our initial hypothesis.

E. Final hypothesis

(11) = (5) reformulated:

(11) a. The affix on a deadjectival verb gives the verb a transitive syntactic primitive along with a causative meaning;

b. Unaffixed verbs are unmarked and consequently likely to have the intransitive, inchoative meaning (adjectives are intransitive).

(11) explains the results provided by the historical data and the uses of affixed and unaffixed verbs. The verb with one of the affixes mentioned above (see par. C.2.) is marked with respect to transitivity, whereas unaffixed verbs, unmarked, can be either intransitive or transitive. This is why 'only transitive' verbs could be found in both morphological classes: some being specifically marked by the affix, some being unmarked with respect to transitivity; and, more precisely, why only unaffixed verbs could be intransitive; we would not expect a marked verb to be uniquely intransitive.

Additional subhypotheses were studied in order to confirm (11), especially where historical data were unavailable.

F. Semantic restrictions as clues to initial syntactic use

Semantic restrictions can be considered to be clues to the initial syntactic use:

(12) a. The more polysemic use of the verb corresponds to its initial syntactic use.

b. The more selectionally restricted use of the verb corresponds to its secondary derived use.

For example, **grandir**, which is clearly initially intransitive (13a), has a figurative meaning associated to the transitive use (13b).

(13) a. il grandit / il grandit à mes yeux/ dans mon
he is growing / he is growing in my regard/ in my

estime
esteem

b. grandir quelqu'un
to make someone look taller (than he really is)

In the same manner, the examples is (14a-c).

(14) a. grossir quelqu'un
to make someone look fatter

b. pâlir quelqu'un
to make someone look paler

c. vieillir quelqu'un
to make someone look older

Embillir, initially transitive, is restricted, in its intransitive derived use, to animate subjects (15b-c), a restriction that does not exist for the object of the transitive use (15a):

(15) a. Jean embellit son jardin / la vérité
John embellishes his garden / the truth

b. ?son jardin / *la vérité embellit
?his garden / *the truth embellishes

c. Marie embellit de jour en jour
Mary grows prettier every day (= animate)

In the same manner:

(16) a. les cheveux éclaircissent au soleil (just a few intransitive uses)
hair lightens in the sun

b. le pull élargit (intransitive: colloquial)
the sweater stretches

c. le temps rafraîchit (intransitive restricted to temperature or weather)
the temperature becomes cooler

d. les jours allongent (intransitive restricted to time)
the days lengthen

G. Intransitive vs reflexive use

Next, intransitive and reflexive use of the verbs were compared, because we suspected these uses could be predicted by the syntactic primitive:

(17) a. If a verb is initially intransitive, the reflexive use, based on the transitive form, will not be synonymous with the intransitive one.

b. If a verb is initially transitive, intransitive and pronominal forms will be synonymous in most of their uses.

This follows from the above claim that the transitive form of a verb which is initially intransitive is more distant from the initial semantics of the adjective. The derivation pattern can be outlined as follows:

(18) a. 1. intransitive ->2. transitive ->3. pronominal

b. 1. transitive ->2. intransitive
->2. pronominal

An example for (17a) and (18a) is (19) where the intransitive use is very different from the reflexive one:

(19 Il grandit ~ il se grandit
he becomes tall ~ he makes himself look taller

The pronominal is clearly different from the intransitive for 26 verbs (out of 34 having intransitive, transitive and pronominal use).

There are 8 to 10 exceptions left like the verbs **jaunir** or **aigrir**: these are listed by Lagane (1967) as symmetrical verbs; i.e. they may have changed classes. They therefore behave like affixed verbs.

(20) is an example for (17b) and (18b); the intransitive use is synonymous with the reflexive one:

(20) le temps refraîchit = le temps se refraichit
the weather becomes cooler

Most affixed verbs - those which have a transitive causative primitive - can have a inchoative reflexive use. In the case where the verb also has an intransitive use, the two uses are synonymous: 19 **a**-affixed verbs, 3 **e**-verbs and 6 **en**-verbs have a reflexive use synonymous with the intransitive one.

The fact that there are synonymous reflexive and intransitive uses in French may indicate an intermediate stage in the development of the language. Recently derived verbs (containing the productive **-is-** or **-ifi-** affix) always use the reflexive for the intransitive, inchoative construction. Some verbs are directly derived as reflexives, as shown by Rothemberg (1974) in her thesis; for example **se miter** (to undergo the process of getting mites [for wool]). She proposes that inchoatives are realized by intransitives only when the subject undergoes a natural (internal) process, whereas they are realized as reflexives when external factors are also required for the process. This could explain the behavior of our exceptions: **jaunir** (to yellow), **aigrir** (to turn sour), etc.

H. Conclusion

In conclusion, the primitive syntactic subcategorization of a deadjectival verb with respect to transitivity can be determined from its morphology. It was shown that the affix of a deadjectival verb is a reliable morphological trace of the initial syntactic use of the verb. Specifically, as I stated in (11), the affix gives the verb a transitive syntactic primitive along with a causative meaning. On the other hand, unaffixed verbs are unmarked both morphologically and semantically and are therefore likely to have the intransitive inchoative meaning.

Once the syntactic primitive is determined by (11), the meaning of the intransitive and reflexive use is, generally, predictable. The intransitive and reflexive uses will be synonymous for transitive verbs, whereas they will be different for initially intransitive verbs.

ACKNOWLEDGEMENTS

This work has been partially subsidized by Research Grant 410-83-0333 from the Social Sciences and Humanities Concil of Canada to Paul Hirschbühler (Syntaxe historique du Français et catégories vides). Special thanks to France Martineau who helped collecting the data.

Christine Kamprath

Morphological Constraints on Epenthesis and Adjunction: Paradigmatic or Cyclic?

In the course of syllabification of a string of segments, it often happens that a consonantal segment is met which cannot be syllabified by the basic syllabification principles of the language, because it violates clustering constraints based on either relative sonority or constituent length.[1] Languages have a variety of ways of dealing with such extrasyllabic segments, as illustrated in (1): a) the extrasyllabic segment may delete, e.g., the word-internal [p] of Spanish esculptura --> escultura; b) it may attach to a flanking syllable, e.g., the initial [s] in the English word 'school'; or c) it may trigger the insertion of an epenthetic vowel with which it can form a separate syllable, e.g., the [s] in the Spanish word 'escuela'.

(1) Alternative Treatments of Extrasyllabic Segments

a) deletion: es.cul.p.tu.ra --> es.cul.tu.ra
'sculpture'

b) attachment: s.kul --> skul
'school'

c) epenthesis: s.cuel.la --> es.cue.la
'school'

The determination of what counts as an extrasyllabic segment is a language-specific matter. Likewise, the choice of how to treat extrasyllabic segments varies from language to language and within a single language.

In the language under consideration here, the dialect of Raeto-Romansh spoken in the village of Berguen, in south-eastern Switzerland, only epenthesis and attachment are employed.[2] The choice between the two depends upon the level of sonority of the extrasyllabic segment, and upon whether it is initial or final. Initially, only extrasyllabic liquids trigger epenthesis and all other extrasyllabic segments attach to the following syllable, e.g.,

(2) Word-initial Extrasyllabicity

/lger/ --> [əlgér](~ [légə]) 'to melt' (~ 'melts)
/rc̀ɛgvə/ -->[ərčɛ́gvə] 'receives'

but

[fner] 'hayshaft' (~ [feŋ] 'hay')
[mner] 'to lead' (~ [méŋə])
[školə] 'school'

In final position, however, not only liquids, but also nasals and [š] trigger epenthesis, and only [s] attaches.

(3) Word-Final Extrasyllabicity

/pevl/ --> [pévəl] 'fodder'
/ezn/ --> [ézən] 'ess'
/pɔlš/ -->[pɔləš] 'thumb'

but

[biəšs] 'animal'

Epenthesis and attachment apply to syllabify extrasyllabic segments as I have just outlined, with one class of exceptions, namely the 1s present tense verb form, in which a final extrasyllabic consonant always attaches, and never triggers epenthesis. 1s verb forms have no vocalic morpheme following the stem and are always consonant-final in regular verbs. Whatever consonants there are after the last stem vowel in the word syllabify as part of the preceding syllable. Thus alternations such as the first and third person singular forms in (4) are to be expected.

(4) Present-Tense Alternations: 1s ~ 3s

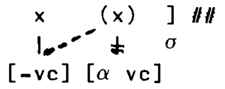
1s	3s	Gloss
tɛŋč	tɛ́ŋǰə	'attain'
pɛkf	pɛ́gvə	'play (a pipe)'
bɛkf	bɛ́gvə	'drink'
kumpfwœ́nt	kumpfwœ́ndə	'confuse'

There is a rule of word-final devoicing, shown in (5), which devoices an obstruent in the rime of the final syllable of a word (followed by a rule spreading devoicing to all tautosyllabic material following it), which may apply if its structural description is met. The devoicing rules in (5) have applied to all of the 1s forms in (4).

(5) a. Final Devoicing
(devoices an obstruent in the rime of the final syllable of a word)

$$[+\text{voice}] \rightarrow [-\text{voice}] / \ \ldots\ x\ \underset{\substack{| \\ [-\text{son}]}}{x}\ (x)\ ]_{\sigma}\#\#$$

b. Devoicing Spread
(spreads devoicing to tautosyllabic material following obstruent)



All first person singular forms observe the pattern shown in (4), even when the final consonant is a nasal or a liquid, that is, a segment which is expected to trigger epenthesis, as shown in (6). The circle under the r̥ or l̥ indicates voicelessness, so these forms have undergone final devoicing. (Unfortunately, final [š] in this position is unattested and thus cannot contribute to this discussion.)

(6) Attachment in 1s Forms with Final Liquid or Nasal

1s	3s	Gloss
dikr̥	dígrə	'endure'
nomn	nómnə	'name'
krekfl̥	krégvlə	'sift'
pɛfl̥	pɛvlə	'feed'

The question we explore here then is this: why does a final extrasyllabic liquid syllabify with the preceding syllable in 1s verb forms but trigger epenthesis elsewhere?

I.e., what motivates the difference between the syllabification of the two forms shown in (7) - the noun [pɛ́vəl] 'fodder' and the 1s verb form [pɛfl̩] 'I feed'?

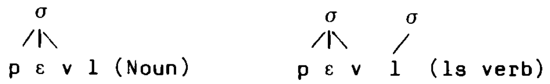
(7) [pɛ́vəl] 'fodder' [pɛfl̩] 'I feed'

There are a variety of conceivable solutions to this question which should be examined not only for their workability, but also for their appropriateness given the nature of the difference to be accounted for. The difference is clearly morphologically based, since only one form among all the forms in the language which have this segmental configuration resists the expected epenthesis solution to extrasyllabicity - and that form is identifiable morphologically as the first person singular present indicative verb form. Therefore, a simple, foolproof solution, and one which reflects the morphological basis for the phenomenon, is simply to assign a lexical diacritic to 1s verb forms to mark their immunity to epenthesis. This solution, however, is in itself unsatisfactory, because it does not identify the nature or source of the aberration. We would prefer a solution which is systematic and explanatory, while also essentially reflecting the morphological basis for the difference between forms like [pɛ́vəl] the noun and [pɛfl̩] the verb.

Two possible systematic solutions are these: a structural solution and a rule ordering solution. The structural solution would motivate a structural difference between the final liquid of /pɛvl/ the noun and that of /pɛvl/ the verb, such that only the former would fit the structural description for epenthesis. This would mean that relative ordering between epenthesis and attachment is irrelevant to the generation of these forms. The ordering solution, on the other hand, depends upon the possibility of ordering epenthesis somewhere in the grammar prior to the formation of 1s verb forms. This would work only if it can be shown that epenthesis does not and indeed cannot apply after 1s verb formation. Let us examine these two solutions in turn, beginning with the structural solution.

The structural solution presented here proposes to motivate a contrast such as that shown in (8). The difference between the two forms is that the final /l/ in the verb form, on the right, has acquired some syllable structure in the course of its derivation, in contrast to the final /l/ of the noun, which is bare of syllable structure until epenthesis applies to make it syllabifiable.

(8)



This solution relies on a clear assessment of the morphological structure of present tense verbs, whose paradigm is shown in (9):

(9) Present Tense Paradigm

pɛfl̥ pɛ́vlən
pɛ́vləs pəvléks
pɛ́vlə pɛ́vlən[3]

How are we to factor out the morphemes here? We may propose an analysis such as presented in (10),

(10) pɛvl + 0 pɛvl + ən
pɛvl + əs pɛvl + eks
pɛvl + ə pɛvl + ən

in which the personal endings attached to the stem are as shown. In this analysis, however, there is no structural difference between the noun, /pɛvl/ --> [pɛ́vəl] 'sieve', and the 1s verb form, both of which would have the underlying structure shown in (11), except for the morphological subscript notation, whether we construe the zero as a null segment or as "nothing".

(11) $[\text{pɛvl} + 0]_N$ $[\text{pɛvl} + 0]_{1sV}$

A second possible morphological breakdown for this paradigm might construe the post-stem [ə] vowel, and the consonants which follow it, as two separate entities, proposing a parallel structure for all the persons, as illustrated in (12).

(12) pɛvl + ə + X pɛvl + ə + n
pɛvl + ə + s pɛvl + ə + eks
pɛvl + ə + 0 pɛvl + ə + n

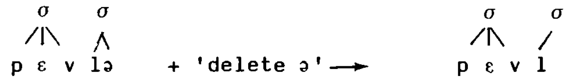
What is the nature of the [ə] intervening between the stem and the personal endings in these forms? It is clearly not the case that these [ə]s are epenthetic, because epenthesis could be motivated only for the 1st and 3rd pl forms, that is, prior to a nasal. If we proposed that the [ə] in the 1st and 3rd pl forms were epenthetic and that the order [ə]s were underlying, we would be proposing a disruption of regularity of morpheme shapes in the paradigm. However, it may be that this [ə] is something like a class vowel, which all the forms in the paradigm get. The apparent problem with the 2pl form, that is, with an excess of vowels in the morphological ending, could be solved by reference to an independently necessary post-lexical vowel coalescence rule which deletes the first of two contiguous vowels, illustrated in (13).

(13) Post-Lexical Vowel Coalescence

otrə ogrə --> otr'ogrə 'other hour'

The advantage of the morphological breakdown presented in (12) is that it gives the 1s form a word-final vowel underlyingly. This vowel is then deleted by the morphological process of 1s formation, represented by the X in (12): whereas the other persons are formed by adding a morpheme (a zero morpheme in the case of the 3s form), the 1s is formed by the deletion of the [ə] vowel. The presence of [ə] prior to deletion, however, makes possible the syllabification of the final consonant as the onset of the syllable of which it is the nucleus, as illustrated in (14).

(14)



Deletion of [ə] then leaves some syllable structure behind on the /l/ in the 1s form, thus distinguishing it from the /l/ of the noun form, which is bare of syllable structure because its excessive sonority in relation to the preceding /v/ prevents it from being syllabified by the basic syllabification rules of the language. The /l/ in the noun meets the structural description for the epenthesis rule, which inserts an epenthetic vowel before an extrasyllabic segment, i.e., before a segment which is left unsyllabified after the application of the basic syllabification rules. But the /l/ in the 1s verb does not meet the structural description for epenthesis, since it is not extrasyllabic, and thus the 1s verb form does not trigger epenthesis.

One of several problems with this solution is that the domain of application of syllabification rules is the word, as has been convincingly argued by Harris (1983), for example. If this is the case, the vowel would be gone before syllabification could apply to it, because the 1s word is not fully formed until all the morphological processes which produce it, including the process of [ə] deletion, have applied. So with this solution we are once again left with the same form of /pɛvl/ as input to syllabification, indistinguishable for both noun and verb. Furthermore, 1s formation by means of deletion is very suspect, even if only because of the asymmetry it introduces into the paradigm, given that all the other persons are formed by affixation. This asymmetry points to more serious problems. [ə] deletion here could not be considered a phonological process, because if it were, it would apply also to the 3s form. But making [ə] deletion a morphological process creates problems of its own. If we propose a 1s "morpheme" which consists not of phonological material but rather the instruction "delete [ə]", which, as a class vowel, is in itself a morpheme of sorts, we have introduced an extremely powerful mechanism into the grammar, unconstrained in principle as to what it could delete. However, this analysis provides the germ of a third possible analysis, namely the notion of an underlying vowel with which a stem-final consonant syllabifies.

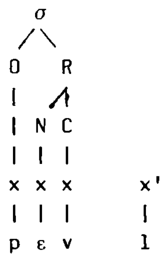
This third possibility for the morphological breakdown of the 1s form, more in line with the affixation process of the rest of the paradigm, proposes the affixation of a morpheme consisting of a somehow phonetically deficient vowel, recognizable by the syllabification rules as vocalic, but deficient enough to warrant its deletion later in the grammar. The morpheme attached could have the form shown in (15).

(15) 1s Morpheme: N
|
x

Thus no deletion is necessary, because this morpheme is affixed. Furthermore, the syllabification rules would know how to identify the slot for syllabification because the syllable nucleus is already present in the morpheme.

By this analysis, then, the structural difference introduced in (8) holds: the /l/ of the noun is extrasyllabic after core syllabification, as illustrated more fully in (16).

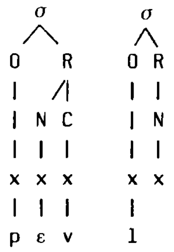
(16) Syllabification of /pɛvl/ → [pɛ́vəl] the Noun



Where x' is an unsyllabified skeletal slot

In the verb, however, the /l/ is initially syllabified as an onset to the ls morpheme, as illustrated in (17),

(17) Syllabification of /pɛvl/ → [pɛfl̥], the ls Verb



Later, the deficient structure constituting the ls morpheme, having no phonetic content, is deleted, but some syllable structure is left behind on the /l/. It is by virtue of this structural detritus that the /l/ of the verb is distinguishable from the /l/ of the noun, which, having never syllabified, is bare of structure. The /l/ of the noun form I will refer to as "stray". The /l/ of the verb form I will refer to as "degenerate", because it constitutes a degenerate syllable. Notice that this use of the term degenerate syllable differs from the use others have made of it (Selkirk, for instance), in that this degenerate syllable comes about only after deletion of syllabic material, whereas Selkirk's degenerate syllable looks like what I am calling a stray, and indeed she does not make a distinction between the two. The epenthesis rule, then, applies only to strays. The attachment rule is less particular. It is the clean-up rule which syllabifies improperly syllabified consonants later in the derivation.

We will assess the merits and defects of this solution after examining the ordering solution.

The ordering solution relies on a version of the level-ordered morphology framework proposed and developed by Kiparsky (1982, 1983), Mohanan (1982), Halle and Mohanan (1985), and others, in which the lexicon consists of ordered strata to which particular morphological processes are assigned. In addition to lexical strata, there is a post-lexical domain, where words concatenate into phrases. Phonological rules are assigned to given morphological strata as their domain of application. Such a framework may motivate the difference between [pɛ́vəl] the noun and [pɛfl̥] the 1s verb form as a difference between the lexical levels on which the two forms are generated. Even if the underlying representations of [pɛ́vəl] and [pɛfl̥] are more complex than meets the eye, that is, for example, if [pɛ́vəl] the noun has some sort of zero morpheme for 'masculine' and if the verb has a similar zero morpheme or a class vowel that deletes, these morphemes will have dissipated before syllabification applies (assuming now that we have no structural difference such as was argued for in the preceding section). Thus if both forms are generated at the same level, they will constitute identical input to either an epenthesis rule or an attachment rule. However, if the two forms are generated on different levels, such that epenthesis applies after noun formation but before 1s verb formation, the former will undergo the epenthesis rule, and the latter will not. Figure (18) presents the lexical organization on which the ordering solution depends: in this analysis nouns are formed on Level One, at the end of which level epenthesis applies, and verbs are formed on Level Two.

(18) Lexical Organization of Berguener-Romansh

Underlying Representation	pεvl) $]_N$	pεvl)X$]_{1sV}$
Derivational Level		
Affixation	p ε v l	NA
Syllabification	p ε v l |/ σ	NA
Epenthesis	pεv əl	NA
Syllabification	p ε v ə l |/ \/ σ σ	NA
Inflectional Level		
Affixation	NA	pεvl
Syllabification	NA	p ε v l |/ σ
Post-Lexical Level		
Phrasal Destressing		
Unstressed V Deletion		
Attachment	NA	p ε v l |// σ

The ordering solution requires that epenthesis applies before the Inflectional Level (Level Two) so that the only recourse the grammar has to syllabifying the extrasyllabic consonants at the end of 1s verb forms is attachment. And, in fact, in this language it is necessary that there be a late, and indeed post-lexical, rule of attachment, because there is an Unstressed Vowel Deletion rule following a Phrasal Destressing rule. The consonants which are left improperly syllabified by this post-lexical vowel deletion are always **attached** to a flanking syllable: they never trigger epenthesis. Therefore we know at least that epenthesis is a lexical rule, because it never applies post-lexically. If it applies just before the inflectional level, [pεvəl] the noun will get an epenthetic vowel, as it should. But all inflected verb forms, and, crucially, first person singular forms such as [pεfl̥], correctly will not get an epenthetic vowel, because epenthetis depends upon prior syllabification, which, as we said, applies only to forms which have the status of words. Since inflected verbs do not undergo affixation and thus do not become words until the inflectional level, they miss out on epenthesis, which is the result we want. Furthermore, there are no verbal inflections which have extrasyllabic final consonants which, in contrast to ls forms, do require epenthesis, so we don't need to order epenthesis after inflection for the proper generation of other verbal desinences either.

There is, however, a problem with this analysis. There are a number of inflected verbs which have an extrasyllabic liquid word-initially, words such as /rδɛgva/ [ərδέgvə] 'receives' (see 19):

(19) /rδɛgva/ → [ərδέgvə] 'receives'

Such forms, like ls forms, cannot undergo epenthesis, because epenthesis applies before they have the word-status that would make them eligible to undergo it. But they apparently do undergo epenthesis, because an epenthetic vowel appears on the surface.

A simple solution to this dilemma is to make two epenthesis rules, one which applies word-initially, and one which applies word-finally. The word-final rule would apply just before the inflectional affixation level, and the word-initial rule would apply just at the end of the inflectional level.

It appears, then, that both solutions work: both the structural solution, under which [pέvəl] the noun and [pɛfl̥] the ls verb are distinguished by a structural difference between stray and degenerate segments; and the ordering solution, under which word-final epenthesis is ordered before verb inflection.

However, each of the two solutions presents negative aspects. The structural solution is questionable because of the fact that nowhere else in this language is it necessary to distinguish between two kinds of improperly syllabified consonants, that is, betwen strays and degenerates. Furthermore, the possibility of leaving behind parts of syllable structure after deletion introduces an extremely powerful mechanism into the grammar. Given this possibility, the grammar could distinguish not only two, but perhaps as many as five or six different kinds of improperly syllabified segments, depending upon how much syllable structure was left behind after deletion. And it would mean that every time a segment was deleted, the amount of structure to be left behind, if any, would have to be specified.

On the other hand, the ordering solution also suffers from several defects. For one thing, it seems that the splitting of the epenthesis rule into two parts many be unwarranted, considering the fact that the two rules are structurally exactly the same, in that they both insert an epenthetic vowel before a stray segment. However, even if we can find a way to have only one epenthesis rule, it is still perhaps a brute-force move to order a rule such as epenthesis deep in the grammer so that an entire class of words, namely all inflected verbs, are immune to it. Epenthesis is not a quirky language-specific rule, but a general principle of syllable-structure well-formedness which should not have to be stipulated by a language in order to apply. Thus it is expected that the generation of grammatical forms should not depend upon a rigid, and perhaps artificial, lexicon-internal ordering of epenthesis.

It may be that the best solution is one which combines the good features of these two analyses and leaves out the bad features. So what are the good features of each of these two solutions? A clear advantage of the structural solution is that it is an essentially morphological solution, because it is engendered by morphological derivation of the 1s verb form. Furthermore, it reflects the fact that the 1s form differs crucially from the others in that it lacks a post-stem vowel where all the others have one. In contrast, the ordering solution seems to be an only apparently morphological solution in that it uses the level ordered organization of the lexicon in order to limit its application, but is not actually tied to morphological form. However, the ordering solution has the advantage of not requiring the introduction of problematical degenerate syllables in order to make the right distinctions.

Let us then propose the combination solution presented in

(20), in which the two lexical levels are collapsed into one, since the crucial step which separated them in (18), namely epenthesis, applies not between lexical levels, but rather after both of them. (The A in /pɛvlA/ represents the abstract vowel shown in (15).

(20) Combination Solution

```
-----------------------------------------------------------
Underlying
Representation
-----------------------------------------------------------
Lexical Level
  Affixation        rčɛgvə            pɛvl          pɛvlA
  Syllab.           r č ɛ g v ə       p ɛ v l       p ɛ v  lA
                      \|/   \/        \|/            \|/ \/
                       σ     σ         σ              σ    σ

  Epenthesis        ə r č ɛ g v ə     p ɛ v ə l     NA
  Syllab.           \/   \|/           \|/  \/      NA
                     σ    σ             σ    σ
-----------------------------------------------------------
Post-Lexical Level
  Delete
    Deficient V          NA                NA       pɛv.l
  Stray Attachm.         NA                NA       pɛvl
                                                    \|//
                                                     σ
-----------------------------------------------------------
```

In this analysis, as in the structural solution, the 1s form is generated by the affixation of a phonetically deficient morpheme, which allows the syllabification of the preceding consonant in onset position, as we saw illustrated above in (17). However, here the epenthesis rule does not apply **before** the inflectional stratum, but just at the end of it. This ordering allows epenthesis to be lexical rather than post-lexical, which, as we saw above, it must be. However, it also allows an epenthetic vowel to be inserted before an extrasyllabic consonant by a single epenthesis rule which applies both word-initially, to a form such as [ər.čɛ́gvə], regardless of whether it is verb or not, and word-finally, to nouns and adjectives such as [pɛ́vəl]. However, it does not apply to 1s verb forms because the consonant in question is not extrasyllabic, but is, rather, syllabified in onset position of the final syllable. Post-lexically, the phonetically deficient 1s morpheme is then deleted, but by a process which is not morphological, but

purely phonological.
Furthermore, in this analysis, when the morpheme deletes, it leaves no syllable structure behind. The extrasyllabic segment left after deletion of the 1s morpheme looks just like the extrasyllabic segment in the noun, which attracted epenthesis. However, since it deletes post-lexically, the only rule which applies to syllabify extrasyllabic segments is the independently motivated attachment rule.

It appears, then, that this solution allows the distinctive treatment of extrasyllabicity in the 1s verb forms to be motivated by virtue of morphological derivation, by means of affixing a phonetically deficient morpheme. Although such a morpheme is a bit abstract, it is motivated by the fact that it reflects the essential difference between the 1s verb form and the other verb forms, namely the absence of a vowel. At the same time, this abstract morpheme has no far-reaching effect on phonological theory, and does not introduce unconstrained possibilities for complexities as does the solution which depends upon left-over syllable structure after deletion of a vowel. Finally, this solution allows both epenthesis and attachment to be late-derivation rules whose unmarked ordering befits their language-universal role in the creation of well-formed syllables.

NOTES

[1] For further information regarding rule-generated syllabification, see Steriade (1982), Harris (1983), and Levin (1985), and references cited therein.

[2] This study is based on data from Lutta (1923) and Cloetta (1964, unpublished) and supported by my fieldwork in the town. For further information see also Kamprath (1985) and Kamprath (1987).

[3] Notice that all these forms are rhizotonic in all persons except the second person singular, which is stressed on the personal ending. The stem vowel of the 2pl. form has undergone the general rule of Unstressed Vowel Reduction, which neutralizes vowels in unstressed position: /i/ and /e/ to [i]; /o/ and /u/ to [u]; and /ɛ/, /a/, and /ɔ/ to /ə/.

REFERENCES

Cloetta, Z.Z. 1964. **Igl Pled da Brauegn.** Unpublished MS.

Halle, M. and K.P. Mohanan. 1985. "Segmental Phonology of Modern English," **Linguistic Inquiry**, 16, 57-116.

Harris, J. 1983. **Syllable Structure and Stress in Spanish: A Nonlinear Analysis.** Linguistic Inquiry Monograph 8, MIT Press, Cambridge, Massachusetts.

Kamprath, C. 1985. "The Syllabification of Consonantal Glides." Paper presented at XIVth North Eastern Linguistic Society meeting, McGill University, Montreal, Canada.

Kamprath, C. 1987. Suprasegmental Structures in a Raeto-Romansh Dialect: A Case Study in Metrical and Lexical Phonology. University of Texas at Austin. Ph.D. Dissertation.

Kiparsky, P. 1982. "Lexical Morphology and Phonology," **Linguistics in the Morning Calm**, Hansin, Seoul, 3-91.

Kiparsky, P. 1983. "Some Consequences of Lexical Phonology", MS, MIT Cambridge, Massachusetts.

Levin, J. 1985. **A Metrical Theory of Syllabicity.** MIT Ph.D. Dissertation.

Lutta, C.M. 1923. **Der Dialekt von Berguen und seine Stellung innerhalb der raetoromanischen Mundarten Graubuendens.** Supplement to **Zeitschrift fuer Romanische Philologie**, #71.

Mohanan, K.P. 1982. **Lexical Phonology.** (MIT Ph.D. Dissertation). Indiana University Linguistics Club, Bloomington.

Selkirk, E. 1981. "Epenthesis and Degenerate Syllables in Cairene Arabic," in H. Borer and J. Aoun, eds. **Theoretical Issues in the Grammar of Semantic Languages.** MIT Working Papers in Linguistics 3.

Steriade, D. 1982. **Greek Prosodies and the Nature of Syllabification Processes.** MIT Ph.D. Dissertation.

Mark Vincent LaPolla

Clitic Movement in Spanish and the Projection Principle

1. Introduction

In this paper we will propose an analysis of Restructuring constructions, in Government-Binding theory, for Spanish and Italian, that does not violate the Projection Principle. We will discuss Rizzi's (1978) analysis and will show how it fails to account for data from Spanish in general and Mexican Spanish in particular. Then we will discuss an alternative movement analysis for Restructuring that builds on Baltin's Landing Site theory (1978, 1982) which unfortunately can only account for a subset of the Restructuring phenomena. Finally we will abstract away from movement and propose a morphological account. This final version of our analysis will also account for auxiliary selection.

2. The Problem

Rizzi (1978) shows that there is a bifurcation in the class of Italian verbs taking infinitival complements. One class of main verbs, i.e. modals, aspectuals and motion verbs, allows the clitic of the embedded verb to attach to either the embedded verb or the matrix verb. The other class, i.e. the rest of the verbs in Italian, allows clitics to attach to the embedded verb only:

(1) a. Piero verrá a parlarti di parapsicologia.
'Piero will come to speak to you about parapsychology.'

b. Piero ti verrá a parlare di parapsicologia.

(2) a. Piero deciderá di parlarti di parapsicologia.
'Piero will decide to speak to you about parapsychology.'

b. *Piero ti deciderá di parlare di parapsicologia.
(p.113)

In "impersonal si" sentences Restructuring verbs (R-verbs) allow the direct object of the embedded clause to become the main subject while the other class does not:

(3) a. Finalmente si comincerá a construire le nuove case popolari.
'Finally PRO will begin to build the new council houses.'

b. Finalmente le nuove case popolari se cominceranno a costruire.

(4) a. Finalmente si otterá di costruire le nuove case popolari.
'Finally PRO will get permission to build the new council houses.'

b. *Finalmente le nuove case popolari si otterranno di costruire.
(p.113)

Also the class of main verbs represented in (1) and (3), which normally take **avere** as their auxiliary, can optionally take **essere** when the embedded verb requires **essere**. The other class, e.g. (2) and (4), cannot:

(5) a. Mario ha voluto tornare a casa.
b. Mario é voluto tornare a casa.
'Mario has/"is" wanted to come back home.'

(6) a. Mario ha promesso di tornare a casa.
b. *Mario é promesso di tornare a casa.
'Mario has promised to come back home.'
(p.114)

Rizzi accounts for these three seemingly unrelated phenomena by proposing a Restructuring rule triggered by just the class of main verbs represented in (1), (3) and (5). The Restructuring rule takes as input a terminal string V_x (Prep) V, where V_x are just those trigger verbs, and outputs some sort of complex verbal predicate V*.

Crucially for Rizzi this process destroys the environment for the Specified Subject Condition (SSC), which blocks clitic movement and object raising in (2) and (4), and allows the embedded verb to subcategorize for the auxiliary. Restructuring accomplishes this by changing an underlying bisentential structure containing a PRO to a simple sentence and, as a by product, eliminates the significance of the PRO in the embedded sentence.

It is crucial for Rizzi that the input to the Restructuring rule be adjacent Vs since "an unstressed pronoun originating in the embedded clause cannot cliticize to the main verb when the whole embedded clause is WH-preposed" (p.120):

(7) a. Questi argomenti, a parlarti dei quali verró al piú presto,...

b. *Questi argomenti, a parlare dei quali ti verró al piú presto,...
(These topics, to talk witn you about which I will come as soon as possible,...)

When only the PP immediately containing the WH-word is preposed Restructuring can occur, since "a parlare dei quali" is not a constituent after Restructuring:

(8) Questi argomenti, dei quali ti verró a parlare al piú presto,...

However, as Rizzi (p. 151-152) points out, there are colloquial styles which marginally admit sentences where Restructuring has occurred but the Vs are not adjacent:

(9) Su questo punto, non ti saprei che dire.
'On this point, I you wouldn't know what to tell.'

Intuitively, Rizzi's proposal of clitic movement depends on destroying the environment for the Specified Subject Condition (SSC).

In Spanish clitic movement is apparently possible across a subject. For example, a clitic can be moved from the embedded verb across a PRO and cliticized onto the next highest verb (Luján 1980):

(10) Quiero poderlo ver__.
(want-I to be able him_i/it_i PRO to see e_i)
'I want to see him (it).'

Furthermore in Mexican Spanish constructions such as (9) are completely acceptable and not marginal. These two facts present a problem for Rizzi's formulation of Restructuring.

In (10), "lo" crosses the Specified Object of "ver", i.e. PRO. "poder" and "ver" are not Restructured since "lo" appears in between them.[2] If (10) had undergone Restructuring then one would expect "lo" to appear before "poder", (11), as it does before "quiero" in (12).

(11) *Quiero lo poder ver__.
*(want-I it_i to be able to see e_i)

(12) Te quiero ver__.
(you_i want-I to see e_i)
'I want to see you.'

With Restructuring formulated as it stands one would have to allow movement into a verbal complex in order to account for (10). This is undesirable. It would violate the structure of the complex verbal predicate, since a verbal predicate is supposed to be of category V. Notice that if one were to move first, and then Restructure, the SSC would be violated, predicting, according to Rizzi, ungrammaticality. It might be argued that this could be gotten around if one assumes the SSC to be a constraint on representation and order it after Restructuring. This, though, is false as the Mexican Spanish example of Inversion in (13) shows us.

(13) Lo quiere Maria tomar__.
(it_i wants Maria to drink e_i)
'Maria wants to drink it.'

If Restructuring occurred before Inversion then the NP "Maria" would violate the complex V. If Inversion occurred first then the structural description for Restructuring would not be met.

(14)[3] a. Yo le sé qué decir__.
(I to him$_i$ know [$_{\bar{S}}$what$_j$ [$_S$PRO to say/
tell e$_i$e$_j$]])
'I know what to say/tell him.'

b. Él sabía darselo.
'He knew to give him it.'

c. Él se lo sabía dar__.
(He to him$_i$ it$_j$ knew to give e$_j$ e$_i$)

(9) and (14a) present problems for the Restructuring analysis because the WH-word "qué" ("che" in Italian) could not move into COMP since there would be no COMP after Restructuring. If the WH-word were first moved into COMP then how could Restructuring apply since the structural description of the rule would not be met? Even if the structural description of the rule were relaxed to allow for WH-phrases, then what would be the status of the WH-phrase and where would it attach in the "complex V"? (Notice that this is also a problem, under Rizzi's analysis, for the optional Prep that he allows in the Structural Description of the rule.) Also if Restructuring were to apply at S Structure then how could the WH-phrase exhibit properties as an operator in COMP at LF? That is how could it bind its trace?

The last problem for Rizzi, and any other Restructuring analysis, is that his analysis violates the Projection Principle which states that subcategorization features must be maintained at DS, SS and LF throughout a derivation. This is not the case if we allow Restructuring. We will account for the problematic data presented above by proposing an analysis that relies on percolation of features and restrictions on that percolation rather than a phrase structure Restructuring rule and restrictions on movement.

Our percolation analysis will allow the "qué" in (14) to remain in COMP and be able to derive sentences like (10) and (13) since it will not rely on bounding conditions of movement or Restructuring.

3. Movement and θ-grids

Spanish has Restructuring. However, it is only apparent in a subset of the cases for Italian. For example, in Spanish the option of choosing "have" or "be" as an auxiliary does not exist. Only "have" can be used as an auxiliary. Also, in this paper we will not look at the impersonal "si" constructions. They would be too lengthy to cover properly here. Therefore the only relevant phenomenon to look at is clitic movement.

Clitic movement in Spanish patterns exactly like clitic movements in Italian with respect to the Restructuring facts:

(16) a. Yo quiero decirle.
b. Yo le quiero decir.
'I want to tell him/her.'

(17) a. Yo decido decirle.
b. *Yo le decido decir.
'I decided to tell him/her.'

Rizzi's account of Restructuring depends on the SSC. That is all the ungrammatical sentences with moved embedded clitics are ruled out via the SSC. Restructuring crucially destroys the significance of the subject of the embedded infinitival, i.e. PRO, for the Restructured strings. Somehow the structure

$[_{vp} [_{s} \text{PRO} [_{vp} \text{V}]]]$

becomes a complex verbal predicate with indeterminate structure but without PRO. Thus the clitic can move without violating the SSC.

In Spanish such sentences like "Quiero poderte$_i$ ver e$_i$" ("I want to be able to see you") are grammatical. "te" has moved from the object clitic position of "ver" to "poder", passing a PRO. This should be a violation of the SSC but the sentence is good.

If Restructuring were to apply to the underlying form of "Quiero poderte$_i$ ver e$_i$", "Quiero poder verte$_i$ e$_i$", one would expect "te" to move to a position either in between "quiero" and "poder" or in front of "quiero", depending on the number of verbs participating in the Restructuring rule.

If movement were to take place before Restructuring then the SSC would be violated. However, if the clitic were base generated in the desired position and then Restructuring applied and the structure checked, post facto, for the SSC violation, the desired results seem to be obtained[4], (18).

(18) DS : Yo$_j$ quiero [$_{VP}$ poderte$_i$ [$_S$ PRO$_j$ [$_{VP}$ ver e$_i$]]].
Restructuring: Yo$_j$ quiero [$_V$ poderte$_i$ ver e$_i$].
SS : Yo$_j$ quiero [$_V$ poderte$_i$ ver e$_i$].

The problem with making the SSC a condition on representation is that this would not allow us to generate such Mexican Spanish sentences as:

"Yo te sé qué decir"
(I to-you$_i$ know what PRO to say e$_i$)

because "qué" would not be in COMP at LF since there would be no COMP at LF. [Notice that making the SSC a condition on representation would be problematic for an analysis that applied restructuring at PF because of (10) and (14a).] That is, (10) and (14a) would still not meet the structural description (SD) of adjacency for Restructuring.

Another solution would be to allow Restructuring to create a Complex Verbal Predicate (CVP), moving the clitic to the left of "poder", and allow the morphology to realize the clitic in the proper place. This could occur at PF if we assume that morphology does not exclusively apply at Deep Structure.

The problem with a CVP analysis is in communicating to the morphology where to realize the clitic. For example, if the VP "quiero poder verte" were Restructured, there would be at least two potential sites, (if not three,) to realize the clitic(s), (Figure 19). How would the morphology know where in the CVP to realize the clitic based solely on the syntax?

(19) a. [$_S$ Yo$_j$ [V [$_{CL}$ e$_i$] [$_V$ te$_i$ quiero
poder ver e$_i$]]]

b. [$_S$ Yo$_j$ [$_V$ [$_{CL}$ e$_i$] [$_V$ quiero
poderte$_i$ ver e$_i$]]]

Another problem for the Restructuring framework is: it is impossible, given the above (morphological) analysis, to account for "Yo le sé qué decir".

In the above constructions clitic movement appears to be allowed only in case the verb to which the raised clitic attaches is a Restructuring verb and all the verbs in between are Restructuring verbs. That is, there is normally one clitic position at the end of a sequence of infinitives but a clitic may optionally attach to a Restructuring verb.

In order to account for Spanish (object) clitic movement in general, and Mexican Spanish clitic movement in specific, a lexical extension of Baltin's (1978, 1982) Landing Site theory is necessary. That is, "landing sites" are specified in the lexicon. For example, clitic movement is obligatory in an AUX Past Participle construction because Past Participles do not have the required "landing site". Therefore a clitic cannot surface on a Past Participle because of the morphology of the Past Participle[5].

In the non-Restructuring cases in Spanish clitics can only appear in one of two places. If the sentence contains only one (finite) verb, the clitic must appear before the verb. If the sentence contains infinitival complements then the clitic must come at the end of the last verb and before any NP. To account for this, one could say that Romance verbal morphology only allows clitics to attach to the left if the verb is tensed and to the right if it is nonfinite.

Clearly this lexical account of clitic sites does not make the right predictions of grammaticality for the sentences in (14). Also one would need a way to constrain the movement in (20) and (21).

(20) a. Yo decido decirle.
(I decide to talk-to-him.)
(I decide to talk to him. (decidir # R-verb))

b. *Yo le decido decir.

(21) a. Quiero decidir verte.
(want-I to decide to see-you)
("querer" = R-verb, "decidir" ≠ R-verb)

b. *Quiero decidirte ver.

c. *Te quiero decidir ver.

The problem presented by the sentences in (20) and (21) is solved if we assume a "transparency condition" that requires a local **(index) chain**, spanning no more than one S node, to relate the moved element to its trace (Baker, personal communication). We will assume that in an unbroken chain all the elements proceed inorder from right to left. The next question is how do we formulate this chain and the restrictions on chain formation.

We can think of a (object) clitic as a spell-out of "non-inherent" features of the noun or verb, following Borer (1984). We would like to propose that inherent features are just those features that affect meaning and argument positions and predicates in Deep Structure. That is inherent features refer in their functional description to A-positions or predicates. For example, inherent features are features like θ-role assignment, feature and categorical selection and subcategorization. Non-inherent features are features like inflection (and perhaps tense) on verbs and number, gender and person on nouns. It is important to note that the non-inherent features which are realized as clitics are distinct **from the inherent features of the noun or verb.** Thus, following Borer, we assume that cliticization does not affect the inherent features of the verb.

Borer proposes a Complement Matching Requirement (CMR) between (object) clitics on the verb and complements of the verb that builds on Stowell's (1981) suggestion that "the assignment of θ-roles to complements by a head can be captured if we assume that a complement transfers a referential index to an available thematic slot in the head" (Borer 1984: 39). The empty slots have to then be filled with referential indices. If the slot cannot be filled, or the matching of θ-role to complement is incorrect, then the derivation is ruled out. The (object) clitic is also contained in the thematic grid of the verb but does not fill a slot, (22). The matching requirement then checks to make sure that the index on the clitic and the one inserted into the θ-grid of the verb match. If they do not match then the derivation is ruled out, (23).

(22) $*[_x\ X, \quad \theta_x]$
$[\ ]$
cl_i

(23) $*[_x\ X, \quad \theta_x]$
$[j]$
cl_i

Borer notes that an (object) clitic is not an argument and cannot therefore satisfy subcategorization requirements. (An empty category (EC) generated in A-position does this. Therefore all clitc constructions are actually clitic doubling constructions.) The structure of the sentence "Juan se la da a Maria" is not (24a) but rather (24b).

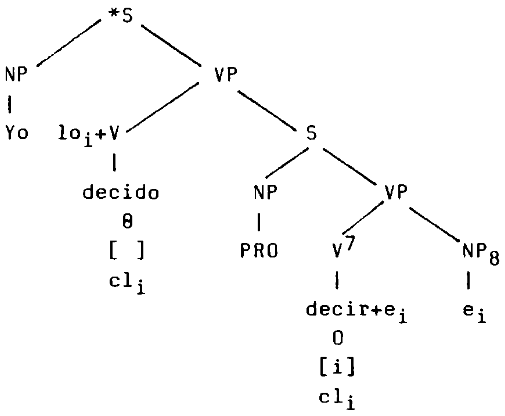
(24) a. Juan se_i la_j da a $Maria_i$.

b. Juan se_i la_j da e_j a $Maria_i$.

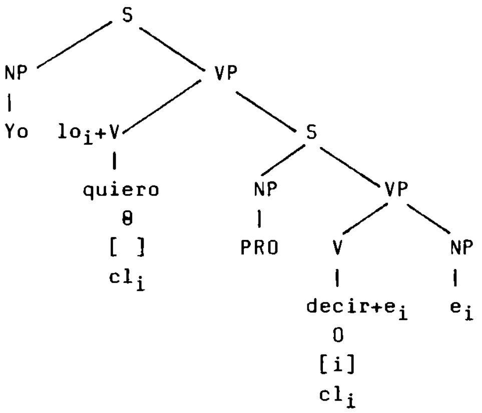
If we assume the above analysis, then the difference between R-verbs and non-R-verbs is simply that non-R-verbs must enter their index into their θ-grids, whenever an object clitic occupies its "landing site", thus forming an (index) chain[6] from the moved element to its landing site. This is optional for R-verbs.

If a clitic lands on a higher non-R-verb, as in (17b), then the CMR will rule out the derivation since that verb's object θ-slot is not filled. (See Figure 25.) If a clitic lands on a R-verb and inserts its index into the R-verb's θ-grid, the derivation, or structure, will be ruled out by the CMR, as for the non-R-verb case. If the index is not inserted then the derivation is good, (26).

(25)


*S
NP
Yo
VP
lo$_i$+V
decido
θ
[ ]
cl$_i$
S
NP
PRO
VP
V^7
decir+e$_i$
0
[i]
cl$_i$
NP$_8$
e$_i$

(26)


S
NP
Yo
VP
lo$_i$+V
quiero
θ
[ ]
cl$_i$
S
NP
PRO
VP
V
decir+e$_i$
0
[i]
cl$_i$
NP
e$_i$

In order for this analysis to work some form of restriction on derivation or structure must be used to disallow movement across more than one S node. That is, we must assume some form of subjacency that will not allow a clitic to "skip over" a non-R-verb on its way to an R-verb.

The movement analysis proposed above will account for clitic movement in Italian and Spanish while avoiding the problems raised above because it does not rely on the SSC but rather the CMR. Therefore this analysis does not have the ordering paradoxes discussed, in Section 1, for sentences with inverted subject, for example (13). However, it does not account for AUX selection alternation in

Italian. A revised analysis will be presented in the next section which will also account for AUX selection.

4. V-grids and Morphological Markings

Intuitively the relationship between origin site and "landing site" of the clitic or AUX and matrix verb, in a "Restructured" sentence, is a type of (long distance) dependency. One way to license this dependency would be to use Bouchard's (1984) notion of V-Chain, used to account for bridge verbs.

Basically, Bouchard's analysis goes as follows: verbs that subcategorize for an $\bar{S}$ complement enter the index of the $\bar{S}$ into their V-grids. This can be done successively creating a chain of θ-indices from the embedded verb to the matrix verb, (27) shown in (28), percolating the indices from one level to the next. Nonbridge verbs break this chain since they do not strictly Subcategorize for an $\bar{S}$ complement.

(27) John thinks that Mary saw_i $Bill_k$.

```
      ┌──────────────────────────────┐
 (28) │  think                       │
      │                              │
      │   ┌────────────┐             │
      │   │ i (saw)    │             │
      │   │            │  = theme    │
      │   │ k = theme  │             │
      │   └------------┘             │
      │                              │
      └──────────────────────────────┘
```

We can use the intuition behind the above account for clitic "movement". R-verbs allow information, i.e. non-inherent features, to be percolated from one level to the next, like bridge verbs which transmit inherent features (Sells; personal communication). The license for transmitting information from one level to the next is different for R-verbs. In bridge verbs subcategorization licenses percolation of information. In R-verbs, there is no license.

Rather than simply saying that R-verbs pass features and non-R-verbs do not, which would put the burden of grammaticality on some poorly formalized aspect of the grammar, a better formalization of license for R-verbs would be to adapt the analysis presented last section. That is, allow all Vs to pass non-inherent features. The distinction then would be that non-R-verbs require that the information they percolate must be used. This would be optional for R-verbs. In other words a non-R-verb must put the index of an object clitic into its θ-grid whenever possible.[9]

If we then assume that the switch from "avere" to "essure" in Restructuring verbs in Italian is a switch in non-inherent feature selection, an account for the AUX switching falls out of what we have discussed.

Whenever percolation of clitic features occurs, all of the non-inherent features, which includes AUX selection, get percolated with the clitic features. Notice that since AUX selection is a change in non-inherent features only, we would have to assume that the Projection Principle only holds for inherent features of a verb and therefore it doesn't hold for clitics, or AUX selection. Therefore we can account for AUX selection with the same mechanisms motivated for clitic movement.

This analysis accounts for the Spanish data and for the Mexican Spanish data, which was problematic for a Restructuring analysis. That is the Restructuring analysis was unable to account for the movement of "le" in "Yo le sé qué decir" because there would be no way to Restructure the above sentence and retain "qué" in COMP (at LF). This analysis can account for the problematic data because it does not rely on changing the structural description of a syntactic configuration but rather on the percolation of non-inherent features and independently motivated principles, i.e. the CMR, to rule out undesired structures.

5. Conclusion

We have shown that Rizzi's formulation of Restructuring phenomena does not account for the facts of Spanish in general and Mexican Spanish in particular. Also Rizzi's account violates the Projection Principle.

We proposed an account that uses percolation of non-inherent features from one constituent level to the next, instead of movement. To do this we assumed, following Borer

(1984) that clitics are non-inherent features and proposed that AUX selection is also a non-inherent feature.

We suggested that non-inherent features freely pass from one constituent level to the next. It was suggested that non-R-verbs must use, if possible, the agreement features that they pass from level to level, at each level, while this was optional for R-verbs. The CMR, an agreement rule, would then rule out the non-R-verb cases since no agreement is possible between a higher verb and a lower object NP which it does not subcategorize for.

We concluded, given our account, that the Projection Principle must only hold for inherent features.

ACKNOWLEDGEMENTS

I would like to thank Denis Bouchard and Marta Luján for all the help, criticism and encouragement they gave me. I would also like to thank Peter Sells for his comments and criticisms. Any mistakes are of course my own.

I would like to thank my informants Patricia Dórame, Ana Santisteban and Marta Luján (who also provided me with many important facts and judgments). I would especially like to thank Patricia Dórame for her kindness.

NOTES

[1] This research was supported by the U.S. Army Research Office under contract DAAG29-84-K-0060. Artificial Intelligence Laboratory, Department of Computer Sciences.

[2] We are indebted to Marta Luján for this argument against Restructuring.

[3] "tener" is another verb which allows this sort of construction. Examples of the forms "tensed-S qué PRO VP" are hard to come by.

[4] Notice that a verbal complex with an embedded clitic could not be base generated since this would require a complex verbal predicate to contain clitics internal to its structure. This would present the same problems that a rule moving a clitic into a complex verbal predicate would present.

[5] Notice that this use of morphology would not refute the above objections to Restructuring as movement and morphology since there are at least two legal "landing sites" possible in a structure like (19).

[6] Intuitively, this concept of index chain is similar to Bouchard's (1984).

[7] The EC cliticized onto "decir" is the trace of the clitic "lo".

[8] This EC is not the clitic trace but rather the object EC discussed above.

[9] If we take the point of view in Bouchard (1984) that clitics are object NP agreement features on the verb, then it becomes clear that the CMR is an agreement rule and the clitic's index just short hand for agreement features.

REFERENCES

Baltin, M. 1978. **Toward a Theory of Movement Rules.** Doctoral dissertation, MIT.

Baltin, M. 1982. A Landing Site Theory of Movement Rules. **Linguistic Inquiry**, Winter 1982, 13(1), 1-38.

Borer, H. 1984. **Parametric Syntax.** Foris Publications.

Bouchard, D. 1984. **On the Content of Empty Categories.** Foris Publications.

Luján, M. 1980. Clitic Promotion and Mood in Spanish Verbal Complements. **Linguistics**, 18(1), 381-484.

Montalbetti, M. 1984. **After Binding: On the Interpretation of Pronouns.** Doctoral dissertation, MIT.

Rizzi, L. 1978. A Restructuring Rule in Italian Syntax. In S. Jay Keyser (Ed.), **Recent Transformational Studies in European Languages.** Cambridge, Massachusetts: The MIT Press.

Zubizarreta, M.L. 1982. **On the Relation of the Lexicon to Syntax.** Doctoral dissertation, MIT.

Claire Lefebvre

Past Participle Agreement in French: Agreement = Case

1. Introduction

Agreement may take place between two elements coindexed under predication (along the lines of Williams 1980 and Couquaux 1981) as exemplified by (1):

(1) a. Subject-tensed verb agreement in the configuration NP_i $INFL_i$;

b. Subject-Predicate agreement in predicative structures of the type **être** NP_i AP_i (assuming Couquaux's configuration (1981)).

c. Subject-Predicate agreement in relative clauses: NP_i $[_{S'}$ Wh_i $[_{S}.....]]$

Agreement under predication consists of a feature copying (e.g. Fauconnier 1974) or checking rule (e.g. Rivas 1977; Lefebvre and Dubuisson 1977) which accounts for the fact that coindexed elements will bear the same features with respect to [α person], [βnumber], [γ gender] and, possibly, [μ Case] (in languages like Latin). Agreement features are realized on [+N] elements (e.g. nouns, adjectives, determiners, and INFL, assuming that in languages exhibiting agreement on the lexical element filling the INFL node, INFL bears the feature [+N] (e.g. Chomsky 1979).

In this paper I will argue that participle agreement in French is not an instantiation of the above type of agreement: it does not operate under predication and it cannot be formulated in terms of a feature copying/checking rule. I will explore the idea that past participle agreement operates from the trace of the NP it governs (section 2) and that past participle agreement phenomena are related to Case in several ways (discussed at length in section 3) summarized in (2):

(2) a. agreement morphology on the past participle spells out the features of the trace of the NP governed by the past participle. Agreement morphology may be viewed as a Case marker on the past participle thus reflecting the Case assigning properties of the past participle.

b. the rule moving the object of the past participle out of its maximal projection (the VP), is an instantiation of Co-Case marking, a special case of co-indexation which Co-Case marks an element moved out of its projections with the projection it is moved out of. In Lefebvre and Muysken (1982), Co-Case marking (or Move CASE) was suggested to account for non-configurational properties of languages. This analysis was shown to account for the rather free word order in languages which have a rich Case morphology. For Romance languages, co-Case marking was suggested to account for cases of extraction of non-heads out of their projections such as quantifier float (e.g. ils sont tous venus), extraction of nominal complements out of their projection (e.g. de qui$_i$ Jean a-t-il vu [la fille [e$_i$]]), and extraposition. Here I suggest that extraction of objects out of VP in passive and ergative constructions (raising from object to subject) is an instantiation of such a rule.

c. agreement morpohology on the past participle participates in a chain linking the past particple, its maximal projection and the NP moved out of this projection (i.e. the preposed object of the past participle. This chain has the properties of a Case Chain such as defined in Chomsky (1981:333 ff).

Predictions of this proposal and some of its consequences will be discussed in sections 4 and 5 respectively.

2. Past Participle Agreement in French operates from the trace of the NP it Governs

This section summarizes the facts of past participle agreement dispersed in the literature and establishes the structural conditions under which past participle agreement takes place. In doing so, I will assume the following:

(3) a. intransitive verbs divide into two groups: true intransitives (e.g. téléphoner) and ergative verbs (e.g. arriver) (Burzio 1981)

b. ergative verbs are characterized by the fact that their surface subject is base generated in the same position as the direct object of a transitive verb: $[_{VP}$ V NP$]$; ergative verbs are distinct from true intransitives, the subject of which is base generated under S: $[_{S}$ NP VP$]$; (Burzio 1981)

c. pronominal verbs (e.g. se donner), inherent reflexives (e.g. se souvenir), verbs occuring with ergative 'se' (e.g. se briser) and passives, are like ergative verbs; their subject is base generated in object NP position.

Contexts in which past participle agreement occurs are contexts involving movement of the direct object out of its projection (VP):

(4) a. les maisons$_i$ que nous avons construites t_i
RELATIVIZED OBJECT

b. ils les$_i$ ont construites t_i
CLITICIZED OBJECT

c. elles$_i$ ont été construites t_i
PASSIVE

d. trois filles$_i$ sont arrivées t_i
ERGATIVE VERB

e. elles$_i$ se sont livrées t_i à la police
PRONOMINAL VERB

f. $elles_i$ se sont souvenues t_i
INHERENT REFLEXIVE

g. la $fenêtre_i$ s'est brisée t_i
ERGATIVE SE

In (4a) the object has moved to COMP. In (4b) the object is cliticezed on the tensed verb. In (4c) - (4g), the object of the past participle has been moved to sentential subject position. Note here that past participle agreement operates independently from the type of trace left by object preposing. According to standard assumptions, the trace in (4a) is a variable, in (4b) it is an anaphor and in (4c) - (4g), a trace of NP.

Contexts which exhibit no agreement are contexts in which the past participle has no object NP, as in (5a) and (5b), contexts in which the NP object position contains lexical material, thus no trace of NP, as in (5c-f).

(5) a. elles ont téléphoné
INTRANSITIVE VERBS

b. il les_i a fait réparer t_i par un ami
FAIRE CONSTRUCTIONS
(Kayne 1984, ex. 1)

c. nous avons construit les maisons
OBJECT IN SITU

d. il est arrivé trois filles

e. il en a repris deux (Kayne 1984, ex. 13)

f. le jour où sont arrivé ces filles

If we assume with McA'Nulty (1983)[1] that past participle agreement operates from the trace of the NP it governs, the facts of (4) and (5) are accounted for in a unified way regardless of the type of auxiliary involved and regardless of the type of past participle involved be it transitive, passive or ergative.

A potential counter-example to the generalization that past participle agreement operates from the trace of the NP it governs is (6) in which the direct object of the past participle has been moved to COMP, thus creating an agreement context.

(6) combien de filles$_i$ est-il arrivé t$_i$
IMPERSONAL IL

In impersonal **il** constructions, however, there is no past participle agreement. I will show below that the facts of (6) constitute only an apparent counter example to the generalization that past participle agreement operates from the trace of the NP it governs.

Assuming, for now, this generalization to be correct, past participle agreement cannot be an instantiation of agreement under predication (cf. section 1) since the past participle and the trace of its object are not in a predicative relationship, although they may be coindexed as a consequence of the Theta assignment properties of the past participle with respect to its object (see Bouchard forthcoming). Moreover, the agreement rule involved cannot be formulated as a feature copying/checking rule of type (7), as suggested in McA'Nulty (1983:147):

(7)
$$\begin{array}{ccc} [+V] & - & t \\ 1 & & \begin{bmatrix}\alpha F\\ \beta F\\ \cdot\\ \cdot\\ \cdot\\ 2\end{bmatrix} \end{array} \longrightarrow \begin{array}{ccc} [+V] & - & t \\ \begin{bmatrix}\alpha F\\ \beta F\\ \cdot\\ \cdot\\ \cdot\\ 2\end{bmatrix} & & \begin{bmatrix}\alpha F\\ \beta F\\ \cdot\\ \cdot\\ \cdot\\ 2\end{bmatrix} \end{array}$$

(= 81) in McA'Nulty 1983)

"(81) is a local rule, so that we expect it not to be an instance of Move α; its structural description need only mention the elements directly involved in the rule, and it requires no indices." (McA'Nulty 1983:147).

The formulation of (7) accounts for the facts of (4); it predicts, however, that in the case of (6), past participle agreement will be grammatical and it is not.

Having established that past participle agreement operates from the trace of the NP governed by the past participle and that it is not a feature copying/checking rule operating under predication, I now turn to my own proposal.

3. Past Participle Agreement and Case

In this section I explore the idea that past participle agreement is related to Case along three dimensions: agreement morphology may be viewed as Case morphology (3.1), object movement out of its maximal projection is an instantiation of Co-Case marking (3.2), elements participating in past participle agreement form a Case Chain (3.3).

The theoretical background relevant to the discussion of each of these points is summarized in (8), (9) and (10). Ergative and passive verbs share the following properties:

(8) a. they assign a theta role to the NP they govern in the configuration $[_{VP}$ V NP$]$

b. they do not assign a theta role to the subject position in the configuration $[_{S}$ NP VP$]$.

c. I will depart from the general assumption that ergative and passive verbs do not have Case assigning properties. Consequently, the motivation for object to subject NP movement, generally assumed to be due to the fact that objects of ergative and passive verbs have to move to subject position in order to be assigned Case, will have to be revised.

Co-Case marking is defined in (9) along the lines of Lefebvre and Muysken (1982):

(9) a. co-Case marking is a special case of co-indexing which co-Case marks elements moved out of their projections with the projections they are moved out of.

b. co-Case markng or move Case is a Wh-type movement rule.

c. since Case is a property of maximal projections only maximal projections (e.g. N''' but not N') are involved in co-Case marking.

d. co-Case marking will only involve elements which bear Case.

Aspects of Case Theory relevant to the discussion are:

(10) a. lexical NPs and variables have to have a Case (Case Filter); traces of NPs do not have Case (Chomsky 1981:175).

b. Case is a property of maximal projections; it can be morphologically realized only on [+N] elements (cf. Lefebvre and Muysken, to appear).

c. Case may be assigned to a chain (cf. Chomsky 1981:333).

d. Case chains are created by Wh-type movement rules.

e. All elements in a chain must bear the same Case (Chomsky 1981, Kayne 1982, Pollock 1983).

f. A chain can contain only one element morphologically marked for Case (Lefebvre and Muysken, to appear).

g. Case chains have two purposes in the grammar:
- escaping the Case Filter
- ensuring Theta connectedness between elements moved out of their projection and the projection they are being moved out of (cf. Lefebvre and Muysken, 1982).

h. Since Case chains are created by Wh-type rules and since they ensure Theta connectedness (cf. Chomsky 1981, Lefebvre and Muysken, 1982) they are relevant at LF.

3.1. Agreement morphology and Case

I propose that agreement morphology on the past participle is the morphological realization of the features [α gender], [β number] and [γ Case] of the trace of the NP governed by the past participle and that agreement morphology may be viewed as a Case marker on the head of a projection, here the past participle. Implicit in this proposal is that ergative and passive verbs have Case assigning properties;

what are the Case assigning properties of passive and ergative verbs? In what sense may agreement morphology on the past participle be related to Case morphology? These questions will be discussed in turn.

I will adopt Pollock's (1983:137) proposal that ergative verbs assign Case to their object and extend this proposal to passive verbs. This proposal is motivated by facts of the type (5f) where the object NP has remained in situ and must receive Case under Government in its basic position. The lack of agreement on the past participle in (5f) and in similar sentences, thus reflecting absence of trace, hence of movement, confirms this analysis. I will disagree with Pollock, however, that the Case assigned by ergative verbs to their object is [+nom], and suggest that ergative and passive verbs in French assign objective Case to their object (as in ergative languages).[2]

I will assume that objective Case has two realizations in French: accusative, the marked form, and what I will refer to as 'ergative', the unmarked form, which is morphologically non-distinct from the nominative form in French. This distinction accounts for the distribution of clitic pronouns in these constructions as illustrated in (11), where both **je** and **me** are analyzed as being objective:

(11) a. je (ergative) suis allée

b. *me (accusative) suis allée

If passive and ergative verbs do indeed assing Case to their object, object preposing is no longer motivated by Case assigning requirements, as has been widely assumed in the literature. What, then, is the motivation for raising? I suggest that object to subject raising, in these constructions, follows from the fact that, in French, the subject position has to be filled, unless the proper context for subject inversion is met (cf. (5f)). This follows from ECP given that French is not a Pro-drop language.

I now turn to the second question raised from relating agreement morphology to Case morphology: in what sense may agreement morphology on the past participle be related to Case morphology? Here two problems arise. First, according to (10b), only [+N] elements are Case carriers: are past participles eligible to bear Case morphology? Here there are two options. Either we can assume that past participles are of the category [+N +V], like adjectives, and consequently they can be Case carriers; or, we can assume with Chomsky (1981) that passive verbs are neutral with respect to the feature [N]; in this option past participle will be allowed, but not required, to bear Case,[3] by virtue of being neutral with respect to the feature [N]. Second, it is clear that the agreement morphology on the past participle does not correspond to the Case assigned to the past participle. That an element may bear the Case of another element is morphologically manifested in languages exhibiting Case floating phenomena. In (12), taken from Quechua, the accusative Case on the nominalized verb corresponds to the Case assigned by the nominalized verb to its object, **warma-**, the latter appearing Caseless at surface structure.

(12)	[warma	riku-sqa-	y	-	ta],	hamu-nqa	(QUECHUA)
	girl	see	NOM	1	AC	come 3FU	

'The girl I saw will come' (from Lefebvre and Muysken, to appear).

Summarizing: like transitive verbs, past participles, passive and ergative verbs assign objective Case to their object; the trace left by object preposing has the features [α gender], [β number], [γ Case]; agreement morphology on the past participle is the morphological realization of these features (which may very well be the way Case absorption is instantiated); the whole features matrix may be seen as Case on the past participle; according to the percolation convention, the features on the head percolate to the maximal projection, in such a way that both the past participle and its maximal projection (VP) will bear the same Case features; conversely, absence of agreement morphology on the past participle will leave the VP node without Case features. In section 4 I return to this point.

3.2. Object preposing and co-Case marking

If we assume on the one hand that passive and ergative verbs are Case assigners (as suggested above) and on the other hand that traces of NPs cannot be in a Case position (as stated in (10a)), the rule which raises object NPs to subject position in ergative and passive constructions cannot be of the type Move NP. In this section I will argue that the rule which moves object NP to subject position in ergative and passive constructions is an instantiation of Move α, where α = Case, a Wh-type rule. In doing so I will be using the standard vocabulary; I nonetheless consider Move Case as co-Case marking (such as defined in (9)) a special case of co-indexation rather than of movement.

Move Case exhibits the diagnostic features of Wh-movement (cf. Chomsky 1977, 1981). Move Case leaves a trace which is in a Case marked position under the assumption that past participles have Case assigning properties; the trace is thus a variable. It operates through a COMP-like position, created by the Case feature on the VP due to percolation of the features on the head encoded by agreement morphology. The landing site of the preposed object is a Theta-less position; in (4a) the object of the past participle is moved to COMP; in (4b) the object is in a clitic position; in (4c) - (4g), the object of the past participle has been moved to subject position, a Theta-less position, since passive and ergative verbs have the property of not assigning a Theta role to their subjects. The landing site for the preposed NP has to be an A' position, Wh-type rules moving elements out of an A position to an A' position; for (4a) and (4b), COMP and clitic positions are both A' positions; is subject position, the landing site for object NP in passive and ergative constructions, an A' position as well? I will assume, with Koopman and Sportiche (1985), that subject position is an A' position, a proposal which, in the case under study, is compatible with the fact that subject position is a Theta-less position. Finally, object to subject raising, in passive and ergative constructions operates in an unbouned way, like Wh movement; it may involve several XP nodes, a phenomenon comparable to unbounded Wh movement involving several S' nodes; in (13) the object of the second past participle has moved out of two XP nodes. (X being equal to V or A depending on the representation of the sentence in (13)).

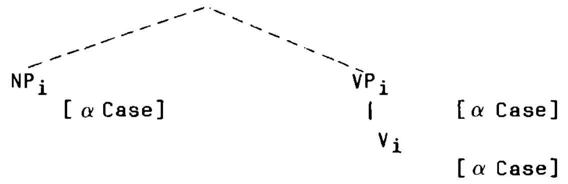
(13) a. la lettre que$_i$ Jean a $_{VPi}$[considérée
$_{VPi}$[écrite e$_i$ par Marie

b. la lettre que$_i$ Jean a $_{VPi}$ [considérée
$_{APi}$[e$_i$ écrite par Marie]]

If this analysis is correct, object preposing in the constructions under study is a Wh-type movement rule, move

Case, not an NP movement type rule. Object preposing may thus be analyzed as an instantiation of Co-Case marking resulting in the co-indexation of the preposed object, the past participle, (due to agreement morphology on it) and its maximal projection through percolation of the features on the head onto the maximal projection. This is illustrated in configuration (14):

(14)



The value of α in (14) is objective; in passive and ergative constructions the NP filling the subject position will thus bear objective Case, a situation which resembles what has been referred to as Quirky Case in other languages. Note here that since subject position is an A' position it will not receive Case from INFL and there will be no Case conflict on the subject.

In the above analysis, Move Case/Co-Case marking exhibits the same diagnostic features as Move Wh. At a more abstract level, however, the two rules may differ, which might be due to several factors: COMP can only have one index assigned to it, while Co-Case marking can occur many times; Wh phrases function as Quantifiers at LF while raised [+N] elements do not (see Lefebvre and Muysken (1982) for further discussion of this point).

3.3. Agreement and Case chains

The co-indexed elements form a chain[4] which has the properties of a Case Chain, properties listed in (10). Move Case defines the trace of the preposed NP as a variable in a Case marked position. The NP moved by Move Case has a Case due to its participation in a chain created by a Wh-type movement rule. The Case Chain is headed by agreement morphology on the past participle which thus functions as an operator at LF. All elements in the chain bear the same Case namely objective Case assigned by the past participle. There is only one element morphologically marked for Case: the past participle. We might say that the Case Chain created by

Move Case and past participle agreement, insures Theta-connectedness between the preposed object and the past participle that assigns it a Theta-role.

In this section I explored the proposal that past participle agreement and related phenomena are linked to Case in several ways: agreement morphology on the past participle may be viewed as Case morphology, extraction of object NPs out of their projection is an instantiation of Move Case or Co-Case marking, Co-indexed elements form a chain which has the properties of a Case Chain.

4. Predictions

The above proposal predicts that past participle agreement will be impossible in impersonal il constructions and in sentences containing the partitive **en**. The analysis in section 3 also predicts agreement facts in sentences referred to as 'the Ruwet (1982) sentences'. I shall discuss these cases in turn.

4.1 Impersonal il constructions

There is no past participle agreement in impersonal il constructions as shown in (15):

(15) a. combien de filles est-il arrivé

b. *combien de filles est-il arrivées

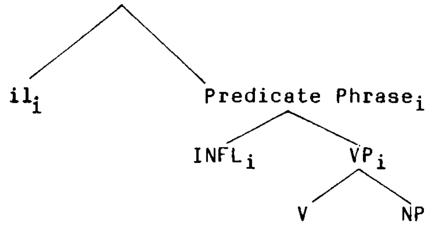
Impersonal il, as it occurs in (15), has the following properties: it is a clitic (Kayne 1975:81-105), it is morphologically marked for number ([-plural]), gender ([-feminine]), person ([3rd person]) and Case ([+Nom]) (Pollock 1983:133). Il does not agree in gender and number with the logical subject of its clause, which follows from the fact that il is not referential. The latter is compatible with the fact that **il** does not occur in a Theta position. In impersonal **il** constructions, the object NP of the ergative verb has to be an indefinite/existential quantifier as illustrated in (16):

(16) a. il est arrivé une fille/*la fille

b. il est arrivé combien de filles/*toutes les
   filles

The data in (16) show that **il** has the property of assigning scope to the direct object of the past participle. Following Williams, a scope marker has to govern - c-command and be adjacent to - the element to which it assigns scope. In (17), the tree structure of (16) **il** c-commands the object NP but it is not adjacent to it. In order to reconcile the French data with Williams' definition of government I suggest that, as a scope marker, **il** governs (c-commands and is adjacent to) the whole predicate phrase it is the subject of (in (17), the tree structure is of the type proposed by Chomsky (forthcoming)).

(17)



The Government relationship is represented in (17) by co-indexation of **il** and the predicate phrase it governs. I will assume that the i index is transferred onto the two nodes dominated by the Predicate phrase node, since both INFL and VP constitute the predicate phrase.[5] Assuming this proposal, **il** will transmit its features to the nodes it is coindexed with in such a way that the VP will bear the feature [+nominative]. We can now explain why there cannot be past participle agreement in impersonal **il** constructions. Agreement morphology on the past participle induces percolations of Case features realized on the past participle to the maximal projection VP. This Case feature is [+objective]. The Case feature transmitted by **il** to the VP is [+nominative]. Past participle agreement creates a Case conflict on the VP; this is why there cannot be past participle agreement in impersonal **il** constructions. In (15a), **combien de filles** will be Co-Case marked with the projection it is moved out of, and will thus be nominative. Lack of agreement on the past participle manifests the impossibility for the past participle to assign Case to its object in this context, a situation which follows from the analysis in section 3.

4.2 En constructions

Past participle agreement is impossible in constructions containing the partitive **en** as shown in (18):

(18) a. Marie en a repris (en=des pommes)

b. *Marie en a reprises

With Milner (1978) and Kayne (1975) I will assume that **en**, in (18), pronominalizes only part of the NP it is part of (probably N') thus not the whole projection. Impossibility of past participle agreement in this context is predicted by the analysis presented in this paper. Past participle agreement is defined as operating from the trace of the NP it governs. According to (9c), the trace must be that of the maximal projection. In (18) the trace governed by the past participle is not that of the maximal projection but that of a N'.

4.3 Ruwet' sentences

Sentences which have come to be known as the Ruwet' sentences such as (19) and (20) exhibit a contrast in past participle agreement.

(19) une femme $_{S'}$[qu$_i$'on a $_{VP}$[dite$_i$ $_{AP}$[t$_i$
belle]]

(20) une femme [qu$_i$'on a dit $_{S'}$[t$_i$[t$_i$ ne pas
être t$_i$ belle]

While in (19) agreement is licensed, in (20) past participle agreement would be ungrammatical. In both cases, a small clause is involved; in (19), SC = AP, while in (20), SC = S'. These data show that while S' is a barrier to agreement, AP is not. These facts follow from the analysis proposed in this paper. Agreement operates from the trace of the NP governed by the past participle and the trace is created by Move Case/co-Case marking; this predicts that past participle agreement will take place only when Move Case is involved.

This prediction is borne out by the data in (19) and (20). In (19) the NP object is moved out of its projection by Move Case hence licensing past participle agreement. In (20) however, Move Case moves the object NP to subject position of the embedded S. From that position the NP is moved to COMP by Wh movement and then from that COMP up to a higher COMP, by Wh movement, not by Move Case. Hence in this context, the past participle **dit** is not in the right configuration to bear agreement features: no NP has been moved out of its projection and hence the past participle cannot participate in a Case Chain headed by agreement morphology on it.

5. Discussion

In this paper I documented a twofold proposal accounting for past participle agreement in French. I showed that the generalization according to which past participle agreement operates from the trace of the object governed by the past participle, accounts in a unified way for past participle agreement phenomena, regardless of the type of auxiliary involved, and regardless of the type of verb involved be they transitive, passive, or ergative. This generalization accounts for the locality of the rule in spite of the fact that, according to my analysis, the 'trigger' for the rule consists in an unbounded phenomenon, object preposing, achieved through Move Case/Co-Case marking. This generalization has advantages over earlier proposals. Burzio's (1981) proposal that "past participle agrees with the element which binds its direct object" is problematic for two reasons: first, it is problematic for impersonal **il** constructions: in (6) it is not **il** that binds the trace, **il** being non-referential, but the Wh phrase in COMP; second as pointed out by Kayne (1985) this formulation does not account for the locality of the phenomenon. The general idea of Kayne's (1985) proposal that past participle agrees with its subject, - an analysis which bears on the assumption that past participles and their object constitute a small clause, and that the object of the past participle is being moved to subject position of the small clause before moving further on - is problematic if we are to take Burzio's thesis on ergative constructions seriously (see Bouchard forthcoming for a thorough discussion of Kayne's proposal).

The bulk of the paper was dedicated to relating past participle agreement phenomena to Case in several ways: agreement morphology on the past participle may be viewed as Case morphology, agreement is triggered by NP preposing, an instantiation of Move Case/Co-Case marking; and finally, the elements involved in past participle agreement and Move Case participate in a chain which has the properties of a Case Chain. This proposal has the advantage of linking NP preposing in the constructions under study to other movement rules in French accounting for extraction of non heads out of their projection (e.g. quantifier float,[6] extraposition, etc. cf. section 1), (a result which is not borne out by Kayne's analysis in terms of small clause, in spite of the fact that the COMP like Case feature on VP in my analysis, shares some properties with Kayne's small clause analysis) and of linking the phenomena under study to word order flexibility in other languages (e.g. Quechua). Another advantage of this proposal is that it allows us to account for all the phenomena related to past participle agreement in a unified and principled way bearing upon the theory of Case (and the theory of movement discussed here in terms of co-Case marking). Absence of agreement in clauses containing the partitive **en** follows from the analysis of agreement as being an instantiation of Case and from the fact that Case is a property of maximal projections. This further provides us with an argument for associating past participle agreement to Case, a property of maximal projections, over an analysis associating past participle agreement to Theta assigning properties of past participles (such as in Bouchard forthcoming), Theta roles being assigned to heads not to maximal projections. It thus appears that it is because of its casual properties that past participle agreement may "constitute a system of clues to the recovery of the thematic structure" (Burzio 1981:57).

Assuming the analysis presented in this paper has some validity, it follows that agreement phenomena in French fall into two types of rules operating in different components of the grammar: agreement achieved through predication, visible at PF (cf. section 1) and agreement resulting from co-Case marking and visible at LF. The first type of agreement falls under Predication Theory; the second type of agreement falls under Case Theory. If this is correct it should not be necessary to consider agreement an independent module of the grammar as proposed in Pollock (1983).

Move Case/Co-Case marking accounts in a unified way for extraction of objects out of their projection, be they Wh, clitics or NPs. If our analysis is correct, it dispenses with NP movement in French. This has consequences for trace theory for, in all cases under study in this article, the trace left by object preposing is a variable in a Case marked position. I leave it an open question, for now, as to whether agreement morphology on the past participle 'absorbs' all of the features of the trace left by object preposing.[7]

NOTES

[*] I would like to thank Denis Bouchard, Isabelle Haik, Monique Lemieux, Yves-Charles Morin, Laurie Tuller, Henk Van Riemsdijk and Edwin Williams for discussing some of the issues presented in this paper with me.

[1] McA'Nully (1983) attributes to Emonds the idea that past participle agreement with **avoir** agrees with the trace of its subject. She generalizes this idea to other constructions. In turn I generalize this idea to all the constructions listed in (4).

[2] See also Sobin (1985) for facts drawn from Ukranian which show that passive verbs assign Case (morphologically realized as accusative) to their object.

[3] If we adopt this second option, we have to revise our definition of Case bearers in terms of elements which do not bear the feature [-N].

[4] For a discussion on agreement chains, see also Pollock (1983).

[5] See also Bouchard (forthcoming) who suggests that impersonal **il** is coindexed with its VP.

[6] In this case also I hold that agreement morphology on the quantifier may be seen as a Case feature, which allows the quantifier to float out of its projection to a Theta-less position (adverbial position). (See Lefebvre, in preparation).

[7] The consequences of the proposal made in this paper will be explored in detail in another paper, given the reduced size this paper must have.

REFERENCES

Bouchard, D. Forthcoming. "A few remarks on past participle agreement."

Burzio, L. 1981. **Intransitive verbs and Italian Auxiliaries.** Ph.D. dissertation, MIT.

Couquaux, D. 1981. "French predication and linguistic theory" in R. May and J. Koster (eds.) **Levels of Syntactic Representation.**

Chomsky, N. 1979. "The Pisa Lectures" (ms.).

Chomsky, N. 1977. "On Wh-movement" in P. Culicover, T. Wasow and A. Akmajian (eds.) **Formal Syntax.** Academic Press.

Chomsky, N. 1981. **Lectures on Government and Binding.** Foris: Dordrecht.

Chomsky, N. Forthcoming. "Barriers".

Fauconnier, G. 1974. **La coréférence: syntaxe ou sémantique?** Seuil: Paris.

Kayne, R. 1975. **French syntax: the transformational cycle.** MIT Press: Cambridge, Mass.

Kayne, R. 1982. "Complex inversion chains in French" in **Wiener linguistische Gazette.** Sonderdruck.

Kayne, R. 1985. "L'accord du participe passé en français et en Italien". **Modèles linguistiques,** VII.1.

Lefebvre, C. (In preparation). "Quantifier float in French".

Lefebvre, C. and C. Dubuisson. 1977. "Les règles d'accord dans la théorie transformationelle: l'accord en Quechua." **Recherches Linguistiques à Montréal,** Vol. 9: 41-90.

Lefebvre, C. and P. Muysken. 1982. "Raising as Move Case". **Linguistic Review** 2:161-210.

Lefebvre, C. and P. Muysken. (To appear). **Mixed Categories: Nominalizations in Quechua.** Reidel, Dordrecht.

McA'Nulty, J. 1983. "Moving features of [E]". Southern California Conference on Romance Linguistics: 107-150.

Milner, S.C. 1978. **De la syntaxe à l'interprétation.** Editions du Seuil, Paris.

Pollock, J.Y. 1983. "Accord, chaines impersonnelles et variables", **Linguisticae Investigationes,** Vol. 7: 131-181.

Rivas, A. 1977. "A theory of clitics". Ph.D. dissertation, MIT.

Ruwet, N. 1982. **Grammaire des insultes et autres études.** Editions du Seuil, Paris.

Sobin, N.J. 1985. "Case assignment in Ukranian morphological passive constructions", **Linguistic Inquiry** 16(4):649-663.

Williams, E. 1983. "Predication", **Linguistic Inquiry** 11(1): 203-238.

Williams, E. 1984. "There-Insertion", **Linguistic Inquiry** 15(1): 131-153.

Juliette Levin

Constraints on Rime-internal Syllabification in French: Eliminating Truncation Rules

0. Introduction

In the past, generalizations over non-surfacing underlying consonants in French have given rise to truncation rules (cf. Schane, 1968; Selkirk, 1972; Dell, 1973), and more recently, to analyses involving consonant-insertion (cf. Tranel, 1981a, 1981b, 1984), unsyllabified (extrasyllabic) consonants (cf. Clements & Keyser, 1983; Clements, 1983), and floating segments unassociated to the skeleton (Hyman, 1984). In this paper, we will attempt to account for certain regular aspects of liaison by following Kahn (1976), Steriade (1982), Levin (1985), and others, in viewing syllabification as an ordered set of structure building rules. Language specific constraints on such rules include relative sonority scales with specified minimal sonority distance (MSD) as well as feature specifications built into the rules of syllabification themselves. Given such constraints, it becomes apparent that the "latent" consonants in French need not be deleted; rather, they fail to be syllabified, and therefore are not part of the phonetically interpreted phonological string. Rather than account for failure of syllabification through extensive lexical marking of extrasyllabicity as suggested in Clements and Keyser (1983) and Clements (1983), or by treating such segments as floating consonants unassociated to the skeleton, we will try to provide evidence pointing to certain regularities which are best accounted for in terms of competing syllabification algorithms.

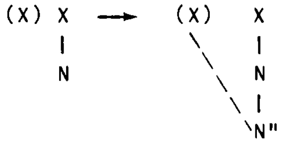
In the first part of this paper we present the facts concerning syllable-initial and syllable-final consonant clusters which motivate a relative sonority scale and rules of N(ucleus) placement, N'-projection, incorporation into both N" and N', and initial and final adjunction to the syllable. In section 2, we review specific phonological properties of latent consonants, and suggest several seemingly significant generalizations which we will aim to capture. After reviewing previous proposals, we move on to an illustration of how one might account for such generalizations without use of truncation rules, insertion rules or widespread use of lexically marked extrasyllabic segments. We argue that two distinct rules of N'-projection can be motivated, one accounting for medial and another for cycle-final consonant syllabification. Facts concerning rime-internal obstruent clusters lead us to propose that the feature structure of assimilated clusters interacts with N'-projection in a non-trivial way.

1. Generating Well-formed Syllables

1.1 Rules of Syllabification

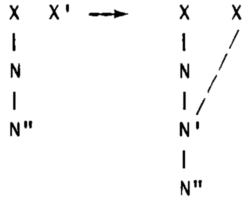
We assume, following Steriade (1982), that syllable structure is a derived property of phonological representations, most accurately represented by an ordered set of structure building rules. Where N(ucleus) is the head of the syllable, N" projection, as shown (1) is universal.[1]

(1) N"-Projection



Projection of N', that is, generation of the subconstituent sometimes referred to as the rime, is a language-particular rule of the type shown in (2):

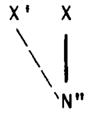
(2) N'-Projection



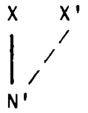
Generation of tautosyllabic pre-nuclear and post-nuclear consonant clusters is accomplished either by rules of incorporation, which are constrained by sonority scales and specified MSD, or, by rules of adjunction, which are not subject to sonority constraints. These rules are illustrated schematically below.

(3) Incorporation

A. Into N"

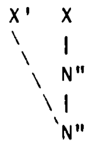


B. Into N'

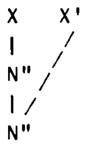


(4) Adjunction

A. Initial



B. Final



As argued by Levin (1985), it appears that within a given language, Incorporation into N" and N' are subject to the same sonority scale, though each rule may require a different specification for minimal sonority distance. We turn now to the attested sequences of consonant clusters in French in an attempt to arrive at the proper formulation of rules of syllabification and language specific constraints on such rules.

1.2 Syllable-initial Clusters

Attested word-initial consonant clusters are given in (5).[2]

(5) French word-initial clusters

A.												
	pt		(kt)							sp,	spl,	spr
	ps		ks		ts					sk,	skl,	skr
	(psk)		(kz)	gz		(dz)				st,		str
			(kv)						(zv)	sf,	sv	
	pn		(kn)	gn						sn		
			(km)		(tm)					sm		
B.												
	pr	br	kr	gr	tr	dr	fr	vr				
	pl	bl	kl	gl			fl	vl	(zl)	sl		
C.												
	pw	bw	kw	gw	tw	dw	fw	vw	zw	sw		
	py	by	ky	gy	ty	dy	fy	vy		sy		

Of these clusters, one can distinguish between those that are obligatorily tautosyllabic and those that are heterosyllabic by the distribution of [e] versus [E]. We assume, following Dell (1973), Tranel (1983) and others that the rule of Closed Syllable Adjustment (CSA) takes /e/ to [E] in closed syllables. CSA is spelled out in (6), followed by examples of the relevant alternations.

(6) Closed Syllable Adjustment

$$e \longrightarrow E \ / \ \underline{\overset{\prime}{\ \ }}\,]\ X_1]_{N'}$$

a. céder	[sede]	cède	[sEd]
b. célébrer	[selebre]	célèbre	[selEbr]
c. élémentaire	[elemãtEr]	élem	[elEm]
d. bénéfice	[benefis]	bénef	[benEf]

While CSA is specified to apply only to stressed vowels, the appearance of obligatory [E] in (7.A), and the absence of surface [E] before OL and OG sequences (7.B,C) leads us to suppose that the rule may apply to certain unstressed vowels as well, providing phonological evidence that such sequences are obligatorily heterosyllabic and tautosyllabic respectively. This syllabification follows that originally proposed by Delattre (1966).[3]

(7) Effects of CSA morpheme-internally

A.	kleptomane	[klEptOman]	lecture	[lEktür]
	epsilon	[EpsilOn]	exacte	[Egzak̇t]
	hebdomadaire	[EbdOmadEr]	excepté	[EksEpte]
	clepsydre	[klEpsidr]	breakfast	[brEkfœst]
	médecin	[mEdsẼ]	flegmatique	[flEgmatik]
	mezzanine	[mEdzanin]	Phlegmon	[flEgmŎ]
	breitschwanz	[brEtšvâts]	ketmie	[kEtmi]
	destin	[dEstẼ]	bélemnite	[belEmnite]
	Chester	[šEstEr]	écheniller	[Ešniye]
	chrestomathie	[krEstOmati]	écheveau	[Ešvo]
	décevoir	[dEsvwar]	lesbien	[lEsbyẼ]
	despote	[dEspOt]	mesmérisme	[mEsmerism]
	espagnol	[Espaɲol]	technique	[tEknik]
	escargot	[Eskargo]		
	espérer	[Espere]		
B.	fébrile	[febril]	beffroi	[befrwa]
			névralgie	[nevralži]
	église	[egliz]	téflon	[teflÕ]
	célébrer	[selebre]	déflation	[deflasyÕ]
	cédrat	[sedra]	février	[fevriye]
	étrange	[etrãž]	échelon	[ešlÕ]
C.	cétoine	[setwan]		
	crétois	[kretwa]		

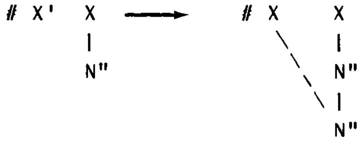
Such data illustrates that well-formed onsets in French consist maximally of obstruent-sonorant clusters, where the sonorant in question is non-nasal. We account for such a constraint by positing the sonority scale shown in (8), with a Minimal Sonority Distance equal to four intervals on the scale (where each natural class defined in (8) is one interval away from the adjacent class).

(8) French sonority scale

[-cons,+cor]	y	
[-cons,-cor]	w	
[+cons,+son,-nas,-lat]	r	
[+cons,+son,-nas,+lat]	l	MSD = 4
[+cons,+son,+nas,+cor]	n	
[+cons,+son,+nas,-cor]	m	
[+cons,-son,+cor]	s,z,š,ž,t,d	
[+cons,-son,-cor]	f,v,p,b,k,g	

Given an MSD of four with only seven notches on the scale, all tautosyllabic syllable initial clusters will be limited to a maximum of two segments, since the second instance of incorporation into N" (3.A) will inevitably violate the MSD. Clusters in (5.A) will all involve instances of initial adjunction since in each case, the MSD is violated. With exception of the rare **psk**, found only in **pschent** [pskEnt]), 'hair-style of Egyptian pharaohs' we can account for the clusters in (5.A) by the simple rule of non-iterative initial adjunction given in (9).

(9) Initial Adjunction (non-iterative)



Notice that this accounts for the sCC clusters as well, since in every case, the consonants following s are the result of incorporation into N", in satisfaction of the MSD.

1.3 Syllable-Final Clusters

Attested word-final consonant clusters are listed in (10).

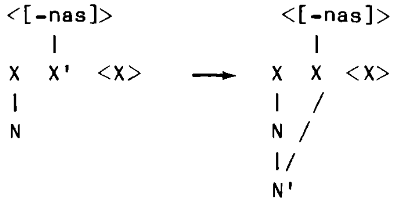
(10) French word-final clusters

A.	pl,bl	fl	gl,kl,rkl	tl	skl,spl	ft,kt,pt,st
	pr,br	fr,vr	gr,kr	tr,dr	str	ps,bs,gs,ks,ts
			gm,km	tm,dm	sm	kst
				(dn,mn)		(dz)
B.	rl					
	rn					
	rm	lm				
	rs,rz,rš,rž,rt,rd	ls,lš,lz,lt,ld	(ms,ns,nd)			
	rf,rv,rk,rg,rp,rb	lf,lv,lk,lg,lp,lb	sp,sk,rsk,lsk			

Given the preliminary sonority scale proposed in (8), the clusters are divided into those which violate this scale (10.A), and those which are consistent with the scale

(10.B). First we discuss the clusters in (10.B). These can be seen to be the result of N'-projection (2), followed by incorporation into N', where incorporation is constrained by a minimal sonority distance of one on the scale in (8) (MSD=1).[4] We assume that the absence of NC clusters is due to the failure of nasals to syllabify word-medially into the coda (N').[5] That is, N'-projection can be formulated as follows:

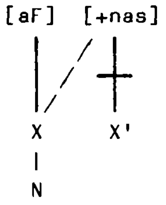
(11) N'-Projection



This rule operates at the word-level, and does have several exceptions (cf. Morin, 1971), including slang terms like **clamse**:[klams], **binse**:[bins], etc.

Post-lexically, unsyllabified nasals are subject to the reassociation rule shown in (12), the result of which are surface nasalized vowels.

(12) Nasal Reassociation (Post-lexical)



The absence of underlying nasal vowels, along with the sonority scale in (8), gives rise to the native speaker's intuitions that the forms in (13) are ill-formed monosyllables, since in all cases the nasalized vowel is derived from an underlying VN sequence followed by /l,r/, which violates the sonority scale, since /l,r/ are more sonorous than the nasals.[6]

(13) Ill-formed tautosyllabic sequences

a. *[pẼr], *[pãr] (from /peNr/, /paNr/)
b. *[pẼl], *[pãl] (from /peNl/, /paNl/)
c. *[pĬy], *[pãy] (from /piNy/, /paNy/)

Deriving surface nasal vowels from underlying sequences of VN is problematic in two respects which we discuss briefly here. First, as pointed out by Tranel (1981a) and Clements (1983), nasalized vowels occur in certain dialects, even though the underlying nasal is syllabified post-lexically under liaison. Examples are given in (14).

(14) Interaction of Nasalization and Liaison

	I. Standard French	II. Québec French (Montréal)
plein emploi	[plE.nã.plwa]	[plẼ.nã.plwa]
pleine envie	[plE.nã.vi]	[plE.nã.vi]

We can account for these differences by ordering rule (12) after post-lexical syllabification (liaison) in dialect I., and before such syllabification in dialect II. Tranel (1984) also points out that the rule of /O/-Raising as stated in (15.a) is also problematic for this account, since Raising appears to be blocked by liaison consonants in some cases (**bon ami** ⟶ [bOnami], *[bonami]), optionally blocked in others (**sot ami** ⟶ [sOtami] ~ [sotami]), and yet obligatorily required with others (**sots ami** ⟶ [sozami], *[sOzami]).[7]

```
(15) a. O  ⟶  o / __ ]        b. O  ⟶  o / __
                                            |
                                            X YZ
                                            |
                                            N]
```

One could reformulate the rule as in (15.b), proposing that it applies post-lexically. Just in case the final consonant is syllabified via liaison, the [O] will no longer be the word-final element on the syllable-plane, but this leaves [sozami] unexplained. The correct generalization seems to be that (15.b) applies lexically (before plural suffixation), and is restricted to cases where the vowel is followed by a non-nasal segment.

We assume for the moment that incorporation into N' is iterative, and that the absence of triconsonantal clusters other than **rsk, lsk**, is in part accidental. However, in proposing a solution to several problems related to liaison, we will posit non-iterative incorporation. Syllabification of clusters like **rsk, lsk** (and even **sp, sk**) will be posited as a secondary result of voice assimilation.

Having accounted for the clusters in (10.B), we now turn to the variety of clusters in (10.A) which are not in conformance with the sonority scale. We assume, following Dell (1973), Anderson (1982) and others, that the clusters in (10.A) which contain final sonorants are the surface result of an optional rule of final schwa-deletion. Under these proposals, the consistent realization of pre-consonantal mid-front unrounded vowels as [E] before such /OLə/ sequences (**règle, célèbre, trèfle, polyèdre, nègre, lèpre, mettre, fièvre**) points either to a more complex statement of CSA (cf. Selkirk's (1978) foot-based account), or to resyllabification following ə-deletion. In light of the success of the sonority scale in (8) in accounting for the majority of surface clusters, we adopt the analysis involving an underlying schwa which is later deleted. This account is strengthened by the fact that there are no attested minimal pairs of the form /reglə/ [reglə]; /regl/ [rEgl], a fact left unaccounted for if all such lexical items are treated as liquid-final.

The remaining final clusters fall into two basic groups: those ending in [s], and those ending in [t]. While certain clusters here are also the result of the optional rule of final schwa-deletion (**dialecte, insecte, syllepse, texte,** etc.), as shown by optional surface schwa phrase-finally, others are clearly not (**infect, suspect, biceps, vertex, est,** etc.). The underlying s-final clusters are amenable to analysis involving final s-adjunction. In addition to this, it appears that we must allow [t]-adjunction in final position, so that final adjunction applies to the class of [+cor,-ant,-voice] segments, though we will revise this proposal shortly.[8]

With this much said, we turn to an analysis of certain phonological properties of liaison, with an eye towards equating the presence or absence of an underlying consonant with the success or failure of that segment to be syllabified.

2. The Phonology of Liaison

The well-known phenomenon of liaison involves zero/consonant alternations in syntactically conditioned phonological environments.[9] Of the seventeen consonantal phonemes in French, (/p,b,t,d,k,g,f,v,s,z,š,ž,m,n,r,l/, only underlying /t,d,s,z,n/ are found to alternate regularly with zero in word-final position followed by a non-vowel initial word, or in phrase-final position.[10] Examples illustrating the C/Ø alternation for each of the underlying segment types participating in liaison are provided in (16).

(16) Liaison alternations

a. /t/:	[t]	petit‿arbre	peti~~t~~ lapin	(petite)
b. /d/:	[t]	grand‿arbre	gran~~d~~ lapin	(grande)
c. /s/:	[z]	gros‿arbre	gro~~s~~ lapin	(grosse)
d. /z/:	[z]	deux‿arbres	deu~~x~~ lapins	(deuxième)
e. /n/:	[n]	bon‿arbre	bo~~n~~ lapin	(bonne)
f. /p/:	[p]	trop‿âpre	tro~~p~~ petit	

As shown in (16), voicing distinctions in obstruents are neutralized in the liaison environment. Anticipating our treatment of latent consonants as unsyllabified (but NOT extrasyllabic) segments, we can state the rule of voice-neutralization as shown in (17).

(17) De/voicing (post-lexical)

$$[-\text{son},\alpha\text{cont}] \dashrightarrow [\alpha\text{voice}]\ /\ \underset{\substack{|\\ X'}}{\underline{\quad\;}}$$

The effect of rule (17) is to limit surfacing latent consonants to [t,z,n,(p)]. Notice that with the exception of /p/, the class of liaison consonants can be specified as the natural class of [+cons,-lat,+cor,+ant] segments.[11] A closer examination of all cases where /p/ alternates with zero in a liaison context points to such examples as lexically isolated forms. A list of p/Ø alternations, including those arising under derivational morphology is provided in (18):

(18) p/Ø alternations

A. Liaison (post-lexical)

beaucoup	beaucoup‿aimé	beaucou~~p~~ trop
trop	trop‿aimé	tro~~p~~ tard

B. Derivational morphology

dra~~p~~	draper, draperie, drapier
galo~~p~~	galoper
siro~~p~~	sirupeux
ca~~mp~~	camper, campement
cha~~mp~~	champêtre
(lou~~p~~	loupiot)

There are several other phonological generalizations with respect to liaison which appear to be significant. First, morpheme-internal latent consonants are preceded in the phonological string either by sonorants (including vowels), or by other latent consonants. There appear to be no cases in which a latent consonant is preceded by a non-latent obstruent morpheme-internally. In (19.A) we see examples of latent consonants preceded by sonorants; in (19.B), we see forms where latent consonants are preceded by other latent consonants.

(19)

A.	B.
peti~~t~~	respe~~ct~~
chau~~d~~	aspe~~ct~~
for~~t~~	pro~~mpt~~

Of particular interest to us is the fact that postvocalic sequences of tauto morphemic stops are subject to a distributional constraint whereby either both obstruents are latent, or both surface. Examples, crucially not involving final-ə, are listed in (20).[12]

(20) Obstruent sequences (Standard French)

	A. Surfacing	B. Latent	C. Unattested
a.	[ps] biceps	(corps)	[...p]~[...ps]
b.	[kt] correct	respect	[...k]~[...kt]
c.	[pt] concept	exempt	[...p]~[...pt]
d.	[st] est	(Jésus Christ)	[...s]~[...st]
e.	[ks] lux	flux	[...k]~[...ks]
f.	[ts] quartz	------	[...t]~[...ts]
g.	[sk] fisc	------	[...s]~[...sk]

As we will see in a moment, the patterning shown above is not adequately accounted for in any proposed analysis: specifically, the absence of alternations such as those listed in (20.C) is treated as accidental.

Notice that while /n/ participates in liaison, /m/ does not. Recall that in medial position, it was suggested that [+nasal] segments can not be syllabified into the rime. As an indication that a different process of syllabification is occuring at word-boundaries, we can look at the forms in (21).

(21) Final Nasals

	A. Surfacing	B. Latent	
a. /m/	islam	nom	(vs. nommer)
	album	parfum	(vs. parfumer)
	harem		
	tram...		
b. /n/	lichen	bon	(vs. bon ami)
	pollen	plein	(vs. plein emploi)
	clown	fin...	(vs. fine)

Clearly nasals can be syllabified word-finally. However, [m]-final sequences are regularly found, as indicated by the continuing list in (21.a), while [n]-final sequences appear to be the exception, rather than the rule. This points to the possibility of singling out the significance of word-final position in terms of rules of syllabification restricted to this phonological environment.

Finally, we take note of several points raised in Tranel's (1984) insightful work. First, when the plural or third-person morphemes are added to a word ending in a latent C, that C is blocked from surfacing phonetically. Second, as remarked on earlier, /O/-raising as stated in (15.a) interacts with liaison in a predictable fashion. Finally, there is a growing class of lexical items with final [p,t,d,n,s,z], where such consonants are always realized phonetically. Examples taken from Tranel are given in (22).

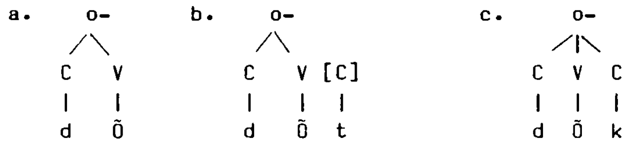
(22) Innovative constant C-finals
a. smoking, titanium, stop, vermouth, club
b. FNAC, OPEP, UNEF
c. prof(esseur), fac(ulté), math(ématiques)

3. A Previous Syllable-based Proposal

Before presenting a proposal to account for the facts seen this far, it will be helpful to quickly review the analysis presented in Clements & Keyser (1983), since it also essentially treats latent consonants as just those consonants which fail to be syllabified, and does away with the need for truncation rules such as those proposed by Schane (1968), Dell (1973) as well as lexically conditioned C-insertion rules of the type proposed by Tranel (1981).

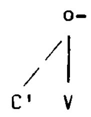
Clemens & Keyser (CK), suggest that truncating obstruents (as well as /r/ in certain cases), are marked in the lexicon by a feature which excludes them from the domain of syllabification rules. Such segments will surface only when they become associated with the syllable node by rule. The rule of liaison for CK then is equivalent to syllabification over word boundary, where the specific rule in question is what we have designated as N"-projection (see (1) above). CK represent the minimally distinct forms **don** 'gift', **dont** 'of which' and **donc** 'therefore' as shown in (23).

(23) Extrasyllabic Consonants (CK; 102)



Given this lexical marking of certain consonants as extrasyllabic ([+EM]), CK are able to account for liaison and truncation by the two rules given in (24).

(24) A. Liaison (CK;104) B. Truncation (CK;105)

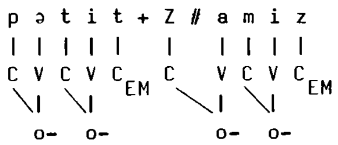


C'---> Ø

(under appropriate syntactic conditions)

CK attempt to restrict the occurence of extrasyllabic consonants to word-final position by paralleling their behaviour to extrametrical elements in metrical theories of stress (cf. Hayes, 1981). Such elements occur only in peripheral position within the domain of metrical structure building rules. Thus, if the domain of syllabification is the word, then extrasyllabic elements should be limited to word-peripheral position. However, as we have seen already, latent consonants do not appear to be limited to word-peripheral position, as illustrated in (25).[13]

(25) Non-peripheral Extrasyllabicity



In some sense then, it appears that the analogy with extrametricality fails. The latent consonants in French act like stray segments in that once they have failed to be syllabified, they may undergo further syllabification, or simply delete. However, their "extrasyllabicity" is clearly not remedied by simple non-peripherality.

Other generalizations proposed by CK concerning the distribution of extrasyllabic segments are listed in (26):

(26) Constraints on extrasyllabic segments (CK;103)
- a. only obstruents and /r/ may be extrasyllabic.
- b. noun and verb stems (root plus derivational endings) never terminate in extrasyllabic consonants.
- c. otherwise, word-final consonants may or may not be extrasyllabic.

Several comments are in order with respect to (26). First, as noted by Schane (1968), [r]~Ø alternations occur only after /E/.[14] In fact, this alternation appears to be limited to the nominal derivational suffixes /-e(r)/ and /-ie(r)/ (**boucher/bouchère; épicier/épicière**, etc.).[15] Furthermore, (26.a) misses what seems to be a significant phonological generalization, namely that with the exception of a handful of words with latent /p/, as noted above, truncated consonants appear to be characterizable as the class of [+cor,-ant] segments. Furthermore, we are left to wonder about the status of final /n/ in a word like **bon** (/bOn/), which is not in conformance with (26.a).

With respect to (26.b), there seems to be some question of descriptive adequacy. Assuming **fruit** to be a root and **fruitier** to be a derived stem, **fruitière** would appear to provide evidence of a latent C in the derivational suffix. If this clause is meant to hold of roots alone, then what of the final segment of /fruit/? While such questions clearly merit further study, we only mean to suggest here that the constraints on extrasyllabic consonants proposed by CK fail to properly limit both the target and structural description of liaison.

4. An Alternative Proposal

4.1 Cycle-final N'-Projection

In an attempt to eliminate CK's seemingly unbridled use of lexical extrasyllabicity, while building on their intuitively appealing proposal that liaison only involves rule (24.A) or our rule (1), and to account for the apparent phonologically significant generalizations summarized in (27), we add to rule (11) of N'-projection the cycle-final rule of N'-projection as stated in (28).[16]

(27) Descriptive statements

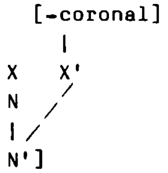
a. /t,d,s,z,n/ are found to regularly alternate with Ø in word-final position; /p,k,m/ alternate with Ø in a small number of lexical items. (16)

b. Nasals appear under N' (in the coda) in word-final position only. (21)

c. Morpheme-internally, latent consonants are preceded by sonorants, or by other latent obstruents. (19)

d. Tautosyllabic tautomorphemic adjacent obstruents agree in voicing. (20.A)

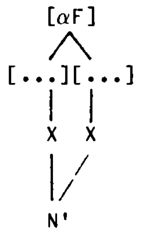
(28) N'-Projection II



As stated, (28) will syllabify a cycle-final segment, if and only if that segment is specified as [-coronal]. As noted above, we achieve this simplification by adopting a theory of underspecification in which /l/ need only be lexically specified as [+lateral], and /y/ need only be specified as [-cons,+hi,-bk (-rd)]. If we compare (28) and (11), we see that in cycle final position, they specify two different processes. Rule (11) allows projection of any segment into the rime, while rule (28) will leave [+coronal] segments unsyllabified. We propose then that on the basis of positive evidence, the language learner is able to determine where it is that rule (28) applies, and where it is that (11) applies.

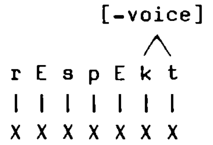
To rules (11) and (28), we add the single condition given in (29) to account for the behaviour of sequences of consonants like those given in (20). Rule (29) states that a stray segment in a feature-sharing relationship with a preceding segment in the Rime is also associated with N'.

(29) Shared-feature Syllabification



The feature in question is [αvoice]. Based on the absence of tautomorphemic tautosyllabic obstruents with distinct voicing specifications, and on evidence of a phonological rule of voice assimilation, we suggest that the pre-syllabification representation of **respect** is as in (30).

(30)



A representation such as that given in (30) could be the result of voice-assimilation, or of a tier-internal instantiation of the Obligatory Contour Principle (McCarthy, 1986), which states that "At the melodic level, adjacent identical elements are prohibited.".

Now the consequences of the conflicting rules (11) and (28) become even more evident. Applying (11), /k/ will be projected from N', and subsequently, (29) will incorporate the /t/ which shares the feature [-voice]. However, the consequence of (29) is directly in conflict with the specifications of rule (28), which does not allow syllabification of the final /t/.

4.2 Summary and Conclusion

Several residual problems are worth noting. First, we are still forced to mark certain consonants as extra syllabic, that is, as exceptions to either rule (11) or rule (28). For example, the final /p/ of /trop/ must be marked as exceptionally failing to syllabify under N', and must thus be marked as [-rule 28]. Likewise, the final nasal in **clown** is syllabified by rule (11), also requiring that it not be subject to (28), or that in this case, unlike the general rule for nasals, (11) takes preference over (28).

Comparing these residual cases of extrasyllabicity to the wholesale use of such a feature within the CK proposal, we see welcome results. First, we are able to limit use of such a feature to consonants in word-peripheral position. Second, we are able to distinguish between the regular latency of coronals as a result of rule (28), and the irregular latency of velars and labials which are marked as exceptions to rule (28). Finally, we are able to account for what appear to be rule-governed constraints on obstruent clusters and their interaction with syllabification without introducing extrasyllabicity.

In conclusion, we have seen that given the sonority scale in (8), rules (11), (28) and (29), and highly restricted use of extrasyllabicity, we are able to account for certain regular phonological properties of liaison without reference to truncation rules, while preserving a unified notion of extrasyllabicity, and capitalizing on what seems to be a fairly unremarkable property, namely that rules of syllabification may distinguish between medial and cycle-final environments. Finally, we have attempted to account for syllabification of obstruent sequences by treating [αvoice] as a shared feature, either as a result of an autosegmental spreading rule, or as a consequence of the OCP at work in lexical representations. This approach requires that rules of syllabification have access to segmental tier-internal structure, as originally suggested by Steriade (1982) for Ancient Greek, and further supported by work on Berber by Dell & Elmedlaoui (1985).

Given that the two rules of N'-projection suggested overlap and conflict in certain cases, it is relevant to ask how such derivations are dealt with both synchronically and diachronically. Synchronically, we have suggested that rule (11) may, in certain cases, take precedence over rule (28). As for the historical progression, few would argue with the claim that rule (28) is slowly making its way into the inner depths of the lexicon, with rule (11) generalizing to cycle-final environments.

NOTES

* Thanks to J.-P. Montreuil, Y. Roberge, and D. Steriade for helpful comments and criticism.

[1] Though we assume that N-placement is also a rule-governed process in French, allowing us to dispense with the distinctive feature [+syllabic], such rules are independent of the arguments at hand. Thus, we start at the post-N-placement level of representation. X' indicates an unsyllabified skeletal slot. See Levin (1985) for a discussion of this notation and N-placement rules in a variety of languages.

[2] Parentheses indicate that the cluster has been found in only a single lexical item. Such items include: **pschent, knout, kmer, xénélasie, kwas, tmèse, dzêta, cténaires, zloty.**

[3] We separate OL from OG for the simple reason that the clusters in (7.C) may not be complex onsets at all, but rather instances where an initial C is followed by a complex nucleus, the first half of which is realized phonetically as a glide. Treatment of CGV sequences under this approach is not uniform, since, given the sonority scale and MSD to be proposed, **mw** is a possible onset, while **rw** is not.

[4] Of the predicted occuring sequences, **Vyw** does not occur. This is a result of the fact that /w/ is not a possible coda in French. At the moment, it is unclear whether to encode this fact by eliminating /w/ from the sonority scale, or by employing a post-syllabification filter. The former approach is argued for in Suzuki's (1986) analysis of Gothic clusters, while the latter approach is adopted in Clements & Keyser (1983).

[5] Note that this approach differs substantially from that of Clements (1983) in that the "extrasyllabicity" of nasals is a rule-governed feature, not an idiosyncratic lexical specification. Nasals are left stray as a result of the regular operation of (11) within the word.

[6] It is critical to this analysis that words such as **genre** contain underlying final schwas, so that the liquid is not tautosyllabic with the preceding nasal. That is, the starred sequences in (13) do not have schwa-final alternants; their behaviour would be identical to words like **par**, **car**, etc.

[7] We do not intend to account for optional raising in phrases like **sot ami** ---> [sOtami] ~ [sotami]. See Tranel (1984) for extensive discussion of these facts.

[8] It should already be apparent that surface word-final [VOL] sequences require a rather unconstrained version of final-adjunction.

[9] See Selkirk (1972) and Manzini (1982) for various descriptions of the relevant syntactic environments for liaison and other syntactically conditioned phonological rules.

[10] Cases of /k/ ~ /Ø/ appear to be limited in modern French to frozen expressions like 'sang et eau' [sãkeo], 'respect humain' [rəspEkümẼ], 'un sang impur' [õesãkẼpü:r]. The absence of liaison in phrases like 'un long entretien' points to the nonproductive nature of k/Ø alternations.

[11] If we adopt underspecified representations, we could further simplify this class to [+cor,+ant], since /l/ could be minimally specified as [+lateral], and /y/ as [-cons,+high,-bk].

[12] This generalization does not extend to Québec French, where alternations such as those shown in (20.C) are attested.

[13] This point is made by Tranel (1984), in exploring the treatment of latent consonants as inter-lexical connective elements.

[14] He also notes that /iy/ goes to [i] and /ul/ to [ü], but that the non-nasal sonorants are not subject to truncation in any other phonological environments.

[15] The other possible instantiation of [r] ~ Ø would be in the /-er/ verb class, should one treat the infinitive as underlying (r)-final, with this /r/ surfacing in future and conditional forms. In stylistic formal speech, surface [r] is possible under liaison, as in 'manger‿à table'.

[16] The significance of (27.d) should become clear momentarily.

REFERENCES

Anderson, S. 1982. "The Analysis of French Schwa," **Language** 58, 534-73.

Clements, G.N. 1983. "A Place for Linking Consonants in French," talk presented at Harvard.

Clements, G.N. and S.J. Keyser. 1983. **CV Phonology**, MIT Press, Cambridge, MA.

Delattre, P. 1966. **Studies in French and Comparative Phonetics**, Mouton, The Hague.

Dell, F. 1973. **Les règles et les sons**, Hermann, Paris.

Dell, F. and M. Elmedlaoui. 1985. "Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber," **Journal of African Languages and Linguistics** 7, 105-130.

Hayes, B. 1981. **A Metrical Theory of Stress Rules**, doctoral dissertation, MIT (reproduced by IULC, Bloomington).

Kahn, D. 1976. **Syllable Based Generalizations in English Phonology**, doctoral dissertation, MIT, Cambridge, MA.

Levin, J. 1985. **A Metrical Theory of Syllabicity**, doctoral dissertation, MIT, Cambridge, MA.

McCarthy, J.J. 1986. "OCP Effects: Gemination and Antigemination," LI, 17.2, 207-263.

Manzini, M.-R. 1981. "Syntactic Conditions on Phonological Rules," ms., MIT, Cambridge.

Morin, Y.-C. 1971. **Computer Experiments in Generative Phonology: Low Level French Phonology**, Natural Language Studies, no.11, Department of Computer and Communication Sciences, University of Michigan, Ann Arbor.

Schane, S. 1968. **French Phonology and Morphology**, MIT Press, Cambridge, MA.

Selkirk, E. 1972. **The Phrasal Phonology of English and French**, doctoral dissertation, MIT, Cambridge.

Selkirk, E. 1978. "The French foot: On the Status of "mute" e," **Studies in French Linguistics**, 1.2, 141-50.

Steriade, D. 1982. **Greek Prosodies and the Nature of Syllabification**, doctoral dissertation, MIT, Cambridge.

Suzuki, S. 1986. "Consonant Clusters, Sonority Scales, and Syllabification in Gothic," ms., University of Texas at Austin.

Tranel, B. 1981a. **Concreteness in Generative Phonology**, University of California Press, Berkeley.

Tranel, B. 1981b. "The Treatment of French Liaison," **Proceedings of the Tenth Symposium on Romance Languages**, H. Contreras and J. Klausenberger eds., University of Washington, Seattle.

Tranel, B. 1983. "On Closed Syllable Adjustment in French," to appear in **Selected Papers from the Thirteenth Symposium on Romance Languages**, L. King and C. Maley eds., John Benjamins, Amsterdam.

Tranel, B. 1984. "French Liaison and Extrasyllabicity," ms., University of California, Irvine.

France Martineau

Variable Deletion of *que* in the spoken French of Ottawa-Hull

1. Introduction

Few studies have devoted complete attention to the deletion of **que** in Canadian French and so far analyses have not reached any agreement concerning the factors conditioning this process. Sankoff, Sarrasin and Cedergren (1971), whose main results appeared in Sankoff (1980 a,b), concluded that **que**-deletion is phonologically conditioned. According to them, the **que**-deletion rule tends to apply more frequently in sibilant environments, less often in other consonantal contexts and still less often in non-consonantal contexts. Connors (1975), in a reanalysis of the data from Sankoff, Sarrasin and Cedergren (1971), concluded, however, that **que**-deletion is conditioned syntactically since the presence of a pronoun following **que** favors its deletion. My work shows that the discrepancy between the preceding studies comes in part from the fact that the **que**-deletion rule involves many factors simultaneously.

2. Methods

I used 30 interviews from oral vernacular conversation. These interviews constitute a sub-part of the corpus of the Ottawa-Hull project directed by Shana Poplack of the University of Ottawa (Poplack, 1983). Like the 120 informants of this project, the 30 informants of my study are evenly divided according to sex, age and neighbourhood. Interviews have been recorded at the informant's residence with quite often the participation of other members of the family. The interviews last two or three hours in general, with, for most of them, many narrations and lively discussions. My corpus amounts to 8201 occurrences. Occurrences of **que** studied appear in complement clauses (1), relative clauses (2), and circumstancial clauses (3).[1]

(1) Je pense **que** / Ø c'est ça.
'I think it's that.'

(2) C'est ça **que** / Ø je dis.
'That's what I say.'

(3) C'est parce **que** / Ø tu veux.
'It's because you want.'

For each factor that has been studied, I calculated the number of **que** deleted out of the total number of **que** that could have been deleted. This was done for each informant. I computed the average frequency of **que**-deletion for the factor studied; I also calculated the standard deviation which gives me a way to evaluate the dispersion of my results. Student-t test was used to verify if the differences found were significant; the confidence interval was .05. Although sociological, stylistic and linguistic factors have been studied in my thesis (Martineau, 1985), only this latter aspect is presented here.

3. Results and Discussion

The total average deletion of **que** is 28%. Complement and circumstancial clauses are significantly more favorable to **que**-deletion than relative clauses; the frequency of **que**-deletion for each type of clause is shown in Table 1. Sankoff (1980b) had already noted that relative clauses favored less **que**-deletion. In relative clauses, **que** is probably less often deleted because, besides linking the subordinate clause to the matrix clause as the conjunction does in complement and circumstancial clauses, it also serves as an antecedent.

	% of que-deletion	Total number of occurrences
Total	28%	8201
Complement clauses	32%	3413
Circumstancial clauses	28%	2722
Relative clauses	22%	2066

Table 1. Total average of **que**-deletion; Average frequency of **que**-deletion according to the type of clause.

3.1 Phonological and grammatical contexts preceding que

The phonological context preceding **que** has been analyzed according to the seven following divisions: vowels, glides, liquids and nasals, non-sibilant fricatives, sibilant fricatives, stops and finally pauses. In complement clauses, sibilants are significantly more favorable to **que**-deletion than vowels and other consonants, except stops. In circumstancial clauses, sibilants are also significantly more favorable than other consonants and vowels. However, in relative clauses, vowels are significantly more favorable to **que**-deletion than consonants, except sibilants (Table 2).

A. **Complement clauses**
   1. Sibilant fricatives (52%; N: 393)
   2. * Stops (41%; N: 160)
   3. Vowels (29%; N: 2272); Liquids and nasals (26%; N: 334); Non-sibilant fricatives (24%; 191); Pauses (24%; N: 62)

B. **Circumstancial clauses**
   1. Sibilant fricatives (32%; N: 2084)
   2. Liquids and nasals (20%; N: 111); Vowels (17%; N: 522)

C. **Relative clauses**
   1. Vowels (26%; N: 1271)
   2. * Sibilant fricatives (19%; N: 196)
   3. Liquids and nasals (17%; N: 387); Stops (12%; N: 125)
   4. Pauses (1%; N: 66)

Table 2. Average frequency of **que**-deletion according to the phonological context preceding **que**.

Each group is significantly distinguished from the others but there are no significant differences within a group. The star that appears sometimes beside a group means that this group is neither significantly distinguished from the group immediately above it nor from the group immediately below it. Absence of a division means that no or less than 25 occurrences have been found. N= total number of occurrences.

Contrary to what Sankoff (1980a,b) concludes, the influence does not appear to be clearly phonological in the context preceding **que**. Important differences among the three types of clauses are found, particularly concerning vowels. However, the influence of certain words in the context preceding **que** seems to be determinant. Since the lexical influence is different from one type of clause to another, they are presented separately.

In the complement clauses, the verb **penser** 'to think' is particularly favorable to **que**-deletion with an average frequency of deletion of 47%. The form **pense** of the verb **penser**, which totalizes 85% of sibilant occurrences preceding **que**, is significantly more favorable to **que**-deletion than sibilant occurrences other than **pense** (**Pense:** 55%; Sibilant occurrences other than **pense:** 35%). Many studies on the deletion of **that** in English point out that the verb **to think** has a favorable influence (McDavid (1964); Bolinger (1972); Wasserman (1976); Kroch and Small (1978)). Likewise, Wanner (1981:51) notes for Italian that: "Within the allowed omissions of **che**, some verbs, rather some privileged forms of some privileged verbs, concentrate the vast majority of such surface deletion on themselves". In Italian, certain forms of the verbs **parere** 'to appear', **sembrare** 'to seem', and **credere** 'to believe' are particularly favorable to deletion of **che**.

Although the verb **penser** is used quite frequently, this aspect does not seem to be the reason of its favorable influence on **que**-deletion since other frequent verbs like **savoir** 'to know' and **vouloir** 'to want' are not very favorable to **que**-deletion (respectively, 17% and 19%). As Kroch and Small (1978) note in their study on the deletion of **that**, lexical formality is often a more important factor than frequency. Use of formal verbs like **croire** 'to believe', which has a meaning similar to **penser**, adds a degree of formality and probably contributes to maintain **que**. The verb **penser** seems to be used as an expression for which the realization without **que** is very generalized. This use could have been strengthened by the fact that there is no change in the mode of the subordinate clause when the verb **penser** is used in the matrix clause (4). This may even lead to an analysis of the subordinate clause as an independent clause (5). Moreover, the verb **penser** seems to be often used as a simple punctuated expression in the discourse, the same way the expression **t'sé** 'you know' is used (6).

(4) Je pense je vais y aller.
'I think I'll go there.'

(5) Je pense: 'Je vais y aller'.
'I think: 'I'll go there'.'

(6) Je vais je pense y aller.
'I will, I think, go there.'

The lexical influence in the complement clauses is not limited to the verb **penser**; verbs such as **sembler** 'to seem' and **falloir** 'to have to' also trigger a high frequency of **que**-deletion (respectively, 52% and 42%) ((7) and (8)).

(7) Me semble il aime ça.
'It seems to me he likes it.'

(8) Faut tu penses à ton avenir.
'You have to think about your future.'

We may contrast the verb **penser**, which does not trigger subjunctive in the subordinate clause and is highly favorable to **que**-deletion with the verb **vouloir**, which does trigger subjunctive but does not favor as much **que**-deletion. However, the mode of the subordinate clause does not seem to have a direct influence on **que**-deletion since we may also find a verb like **falloir**, which even if it triggers subjunctive, is highly favorable to **que**-deletion.

In the circumstancial clauses, sibilants preceding **que** are also highly favorable to **que**-deletion. However, this favorable character of sibilant consonants is essentially caused by, what I will call, the complex conjunction **parce que** which totalizes 95% of sibilant occurrences preceding **que**. It is obviously difficult in this context to distinguish the influence of **parce que** from that of the sibilant consonants. But, if the sibilant occurrences linked to **parce que** are removed from the computation, the remaining sibilant consonants are neither more nor less favorable to **que**-deletion than other consonants and vowels. Therefore, **parce que**, which appears in 73% of circumstancial clauses, has a determinant influence on **que**-deletion; the relatively high frequency of **que**-deletion that **parce que** triggers (32%) explains most of the results in the circumstancial clauses. The lexical influence is overwhelming in circumstancial clauses; for instance, as shown in (9), among temporal circumstancial clauses, there is a great difference in **que**-deletion frequency according to the complex conjunction used.

(9) Depuis que (33%); après que (16%); aussitôt
que (14%); avant que (11%); jusqu'à temps que (7%).

The complex conjunction **parce que** has a special status compared to the other complex conjunctions in (9) because it is constituted by a preposition (**par**) which has incorporated the sequence **ce que**. A similar process has occurred for the complex conjunction **lorsque**.[2] In fact, frequent use of **parce que** in addition to high frequency of **que**-deletion for this complex conjunction may have contributed to a reanalysis in terms of lexicalization for the form **parce**.

In the relative clauses, vowels preceding **que** are significantly more favorable to **que**-deletion but, contrary to what was found in the two other types of clauses, this influence is not clearly linked to a particular word. However, pronouns are significantly more favorable to **que**-deletion than other words preceding **que** in relative clauses (Pronouns: 32%; Other words: 19%). Moreover, pronouns finishing with a vowel like **moi** or **ça** constitute around one third of the words that appear in the vocalic context preceding **que** (for instance, C'est ça que j'aime) and they are significantly more favorable to **que**-deletion than other words finishing with a vowel (Pronouns finishing with a vowel: 33%; Other words finishing with a vowel: 22%). This latter group is not even significantly more favorable to **que**-deletion than consonants. Although pronouns may be linked to a high frequency of **que**-deletion, it seems that this result may perhaps be simply an effect of the fact that they often appear in cleft relatives (92% of pronouns preceding **que** appear in cleft relatives). Cleft relatives (10) are significantly more favorable to **que**-deletion than non-cleft relatives (11) (Cleft relatives: 26%; Non-cleft relatives: 15%).

(10) C'est ça que je dis.
'That's what I say.'

(11) Je regarde les vues que mon mari aime.
'I watch the movies that my husband likes.'

Most of the cleft relatives belong to the type **c'est..que**, as (10) exemplified. The first part of the structure has an ascending intonation while the second part has a descending intonation. The ascending intonation preceding **que** which is created by the particular structure of cleft relatives has probably an influence on **que**-deletion. In fact, Connors (1975) notes that the accentuation preceding **que** favors **que**-deletion. Moreover, Bolinger (1972) makes the same observation for deletion of **that** in English. In brief, in relative clauses, the particularly favorable character of vowels and pronouns is probably an effect of cleft sentences.

3.2 Phonological and grammatical contexts following que

In the context following **que**, the phonological context is an important factor of the conditioning. The phonological context has been divided the same way as the one preceding **que**. In the three types of clauses, there is a tendency for stops and sibilant and non-sibilant fricatives to be the most favorable to **que**-deletion, followed by liquids and nasals, these being followed by vowels (Table 3). In fact, the phonological context following **que** is the only factor among all the factors studied for which the tendencies are similar from one type of clause to another.

A. **Complement clauses**
  1. Stops (56%; N: 440); Sibilant fricatives (58%; N: 1262); Non-sibilant fricatives (57%; N: 32)
  2. Liquids and nasals (17%; N: 300)
  3. Pauses (6%; N: 150); Vowels (6%; N: 1229)

B. **Circumstancial clauses**
  1. Sibilant fricatives (49%; N: 826); Non-sibilant fricatives (50%; N: 37); Stops (41%; N: 210)
  2. Liquids and nasals (22%; N: 456); Vowels (15%; N: 983)
  3. Pauses (6%; N: 210)

C. **Relative clauses**
  1. Sibilant fricatives (49%; N: 797); Non-sibilant fricatives (58%; N: 30)
  2. Stops (28%; N: 130)
  3. Liquids and nasals (12%; N: 90)
  4. * Pauses (7%; N: 32)
  5. Vowels (1%; N: 987)

Table 3. Average frequency of **que**-deletion according to the phonological context following **que**.
See legend of Table 2.

The non-sonorant consonants (obstruents) are more favorable to **que**-deletion than the sonorant elements (liquids, nasals, and vowels); the sonorant consonants (liquids and nasals) are more favorable to **que**-deletion than vowels. Therefore, the relevant factor is the sonority of the segment following **que**. As was pointed out by Sankoff (1980a), sibilants following **que** favor **que**-deletion, except that it is not only sibilant consonants but more generally the less sonorant consonants which favor **que**-deletion. Finally, pauses do not favor deletion of **que**, probably because they cause the schwa to be maintained, hence also the consonant [k].

According to Connors (1975), the fact that sibilants are favorable to **que**-deletion is caused by the pronouns **ça**, **ce**, and **je**, the consonant of which is sibilant. Also, according to her, pronouns following **que** are in general more favorable to **que**-deletion than other words and therefore, **que**-deletion rule is syntactically conditioned.

In my work, pronouns are significantly more favorable to **que**-deletion, but only in complement clauses. Even there, the influence is not only syntactic since, after pronouns following **que** are removed from the computation, phonological differences remain for the other words following **que**. It is difficult to explain the relative influence of pronouns and the explanations proposed until now are not really satisfying. For instance, many studies on deletion of **that** have noted the favorable character of the pronouns (McDavid (1964); Bolinger (1972); Wasserman (1976)). According to these, the fact that pronouns better indicate the beginning of the subordinate clause than other words when the complementizer is absent could be the reason for this difference. However, this explanation is not fully adequate because the flow of the spoken language should be rapid enough for the sentence structure to be analyzed appropriately. Another possible explanation, frequent use of pronouns, cannot be maintained since pronouns are also extremely frequent in relative and circumstancial clauses without being particularly favorable to **que**-deletion.

3.3 Discursive aspect

The influence of the complexity of the sentence has been studied by counting the number of clauses preceding and following **que**. I counted any clause which is not independent and which is therefore either a subordinate or a matrix clause.

There is no significant difference according to the number of clauses following **que** for the three types of clauses. However, some significant differences have been found according to the number of clauses preceding **que** in complement and relative clauses but not in circumstancial clauses (Table 4).

A. **Complement clauses**
Less than 2 clauses: 34%; N: 3013
2 clauses or more: 21%; N: 400

B. **Circumstancial clauses**
Less than 2 clauses: 27%; N: 2404
2 clauses or more: 27%; N: 318

C. **Relative clauses**
Less than 2 clauses: 23%; N: 1887
2 clauses or more: 9%; N: 179

Table 4. Average frequency of **que**-deletion according to the number of clauses preceding **que**.
N= Total number of occurrences.

The presence of two clauses or more preceding **que** contributes to maintain **que** in complement and relative clauses. In a complex sentence, **que** is more often kept, probably so that the clauses are better delimited. As Wasserman (1976:149) remarks for English: "Use of the **that**-complementizer is sometimes reputed to act like a traffic rule for speech processing". We may suppose that **que** has the same role in French in complement and relative clauses. However, in circumstancial clauses, this factor does not have an influence. The loss of **que** has not the same impact as in relative and complement clauses since there remains the first term of the complex conjunction (for example, **parce** from **parce que**) which serves to indicate the beginning of the subordinate clause (ex., C'est parce je veux).

4. Conclusion

In conclusion, the frequency of deletion of **que** is quite high and reflects the fact the **que** is most importantly an element which serves to link clauses. **Que**-deletion is not a new phenomenon in French since it already existed during the Middle Ages. Deletion of the complementizer in medieval Italian is also attested by Wanner (1981). As he remarks, nowadays, complementizer deletion is a rule shared among others by English, German and, in a less productive way, by Italian and Spanish. However, most grammarians consider that **que**-deletion no longer pertains to Modern French. It is nonetheless still present in spoken Canadian French.

My work shows clearly the need to take many factors simultaneously into consideration in order to give a complete account of **que**-deletion. **Que**-deletion is without any doubt influenced by the phonological context following **que** since it is the only factor for which similar tendencies appear systematically from one type of clause to another. The less sonorant the element following **que** is, the more **que**-deletion is favored. **Que**-deletion would thus help to avoid the creation of a consonantal cluster. Presence or absence of consonant [k] of **que** is partly determined by the possibility of forming an onset with the following segment.

Que-deletion is also influenced by other factors. First, relative clauses are less favorable to **que**-deletion, probably due to the semantic content of **que**. Second, in the phonological context preceding **que**, the results differ a great deal from one type of clause to another and there is no phonological influence completely independent from a lexical or syntactic influence. Finally, the number of clauses preceding **que** has an influence on **que**-deletion only in complement and relative clauses; this seems to indicate that the presence of **que** is necessary when the comprehension of the sentence might be delayed by its absence.

NOTES

[1] Occurrences of **que** that do not appear in these three types of clauses have not been studied. Structures which have not been analyzed in the present study include, first, cases for which no instances of **que**-deletion have been found, as in comparative clauses (ex., Elle est moins grande **que** toi). Second, alternations between **ce que** and **qu'est-ce que** have not been studied. Third, **que** must have introduced a clause to be part of the corpus; for this reason, cases such as **rien que** have not been studied. Fourth, cases of **que**-insertion (ex., **quand que, comment que, pourquoi que**) have not been analyzed. Finally, ambiguous cases have been rejected; occurrences of **pis que** raise a problem because it is often impossible to determine if **pis** coordinates or subordinates the following clause (Lefebvre and Fournier, 1978), and therefore, they have been eliminated.

[2] According to Lefebvre (1982), the sequence **ce que** /sk/ is an allomorph of the complementizer **que**. For her, the allomorph **ce que** is found in Canadian French in complex WH-phrases such as **comment-ce que** /k mask/ and **pourquoi-ce que** /purkwask/, these forms having been lexicalized. The /k/ of the cluster /sk/ can be deleted by the rule of reduction of final consonantal clusters.

REFERENCES

Bolinger, D. 1972. That's that. The Hague: Mouton.

Connors, K. 1975. L'effacement de QUE-règle syntaxique. Recherches Linguistiques à Montréal 4.17-33.

Kroch, A. & C. Small. 1978. Grammatical Ideology and Its Effect on Speech. Linguistic Variation: Models and Methods, ed. by David Sankoff, 45-55. New York: Academic Press.

Lefebvre, C. 1982. Le répertoire des mots WH en français vernaculaire et leur insertion dans la grammaire du français. La syntaxe comparée du français standard et populaire: approches formelle et fonctionnelle, ed. by Claire Lefebvre, vol. 1, 73-125. Quebec: Office de la langue française.

Lefebvre, C. & R. Fournier. 1978. Les relatives en français de Montréal. Cahiers de linguistique 8.273-294.

McDavid, V. 1964. The Alternation of "That" and Zero in Noun Clauses. American Speech 39.102-113.

Martineau, F. 1985. L'élision variable de (que) dans le parler d'Ottawa-Hull. University of Ottawa M.A. thesis.

Poplack, S. 1983. The care and handling of a mega-corpus: The Ottawa-Hull French Project. Proceedings of NWAVE XI, ed. by Ralph W. Fasold & Deborah Schiffrin. Washington: Georgetown University Press. (to appear).

Sankoff, G. 1980a. A Quantitative Paradigm for the Study of Communicative Competence. The Social Life of Language, ed. by Gillian Sankoff, 47-79. Philadelphia: University of Pennsylvania Press.

Sankoff, G. 1980b. Above and Beyond Phonology in Variable Rules. The Social Life of Language, ed. by Gillian Sankoff, 81-93. Philadelphia: University of Pennsylvania Press.

Sankoff, G., R. Sarrasin, & H. Cedergren. 1971. Quelques considérations sur la distribution de la variable **que** dans le français de Montréal. Paper read at the annual meeting of the Association canadienne-française pour l'avancement des sciences.

Wanner, D. 1981. Surface Complementizer Deletion: Italian che-Ø. Journal of Italian Linguistics 6.47-82.

Wasserman, R. 1976. Theories of Linguistic Variation. Ph.D. dissertation reproduced by Indiana University Linguistics Club.

Donna Jo Napoli

On Predication and Identity within NP's

There are at least two major approaches to predication theory in recent work within GB.

One approach takes subjects and predicates to stand in the relationship of sister nodes that form a constituent together (perhaps with some adjuncts or adverbials as additional sisters within that constituent) at SS (see LGB, among others). With this analysis the subject and predicate mutually c-command each other. The debates with this approach center around what category this constituent is (S, AP, PP, ...) in a given construction.

A second approach takes subjects to c-command and to be external to the maximal projection of their predicates with the predicates c-subjacent to their subjects but with no stipulation that the subject and the predicate form a constituent, all of these conditions holding at SS (see Williams 1980, among others). A variation on this general approach takes subjects and predicates to be sister nodes, that is, to mutually c-command each other, but again not necessarily to form a constituent (see Schein 1982). I lump these two approaches together since both of them take the theta criterion as limiting any given NP to receiving at most one theta role from any single predicate, but they allow an NP to have more than one theta role if it is an argument of more than one predicate.

Below, I analyze an instance of predication for which both approaches above are inadequate with respect to the configurational demands they make on subjects and predicates and for which the theta criterion must be taken as Williams and Schein propose it. Specifically, in the construction analyzed here, the subject argument of the predicate is not external to the maximal projection of the predicate. Also, given a particular analysis of the external structure of NPs that are complements to N (where the relevant question is whether such NPs are inside PPs or are simple NPs with a Case marker that is homophonous to a P), the subject argument of the predicate may not c-command that predicate. Furthermore, the subject argument of the predicate may receive a theta role both from the predicate in question and from some other predicate.

The analysis of the construction here is consistent with the theory of predication in Napoli 1987. In this theory there are two kinds of predication: primary and secondary. Primary predication is between the first XP sister to INFL (assuming the X Bar theory of Chomsky 1986) and its structural subject (that is, [NP,S]). Secondary predication is all instances of predication other than primary predication. In certain structures the subject argument of a secondary predicate must receive a theta role from some other source, as well as from the secondary predicate. For our purposes the relevant point is that in this theory the only configurational restriction on all predication is that in SS a subject argument must not be separated by any of its barriers (in the sense of Chomsky 1986) from the lexical head of its predicate.

Throughout this paper the term subject is used in two ways. A syntactic subject is the structural entity [NP,S]. A subject argument is that entity specially marked in the lexical structure of some item as its subject argument. For N the subject argument is "that argument by virtue of which the N is used predicatively when it appears in predicative position, and by virtue of which it is used referentially when it appears in referential position...". (This is a quote from Williams 1984b, p.613, who is writing with respect to external argument. I have thus applied his definition to the concept of subject argument.) For PP the subject argument is that argument by virtue of which the PP is used predicatively when it appears in predicative position.

This paper is organized as follows. In section 1 I show that phrases like **quel matto di Giorgio** 'that madman George' are externally NPs. In section 2 I show that the internal breakdown of these NPs has **matto** as the head N and **di Giorgio** as a PP sister to the head N. Thus **Giorgio** neither c-commands nor is external to the maximal projection of **matto**. In section 3 I show that **matto** is a predicate taking **Giorgio** as its subject argument, where the initial specifier on the whole NP specifies the whole NP. Section 4 concludes with some remarks about the consequences of the analysis given here for grammatical theory.

1. The External Structure of Phrases like quel matto di Giorgio

I argue here that the external structure of these phrases is NP. In doing this, I discuss semantic properties these phrases have as well as syntactic properties.

First, they occur in all and only those positions that NPs can occur in.

(1) Quel matto di Giorgio mise il libro nell'armadio.
'That madman George put the book in the wardrobe.'

(2) Avvicinai quel matto di Giorgio alla finestra.
'I brought that madman George to the window.'

(3) Hanno mandato la lettera a quel matto di Giorgio.
'They sent the letter to that madman George.'

(4) Oreste ha organizzato con quel matto di Giorgio di andare al cinema.
'Oreste arranged with that madman George to go to the movies.'

(5) Tuo fratello, quel matto di Giorgio, assomiglia ad un grillo.
'Your brother, that madman George, resembles a cricket.'

Second, they can undergo NP movement, as in raising, passive, ergative constructions, si constructions, and inversion.

(6) Quel fesso di Paolo sembra lamentarsene.
'That dumbo Paul seems to be complaining about it.'

(7) Quel fesso di Paolo è sprezzato da tutti.
'That dumbo Paul is discounted by everyone.'

(8) Quel fesso di Paolo è arrivato.
'That dumbo Paul has arrived.'

(9) Quel fesso di Paolo non si vede più.
'That dumbo Paul isn't seen anymore.'

(10) L'ha fatto quel fesso di Paolo.
'That dumbo Paul did it.'

Third, they trigger subject-verb agreement.

(11) Quei coglioni dei suoi fratelli non ci vengono/
pl. pl.

*viene.
sg.

'Those jerks of her brothers aren't coming.'

Fourth, they are non-propositional. Thus, they cannot bear a propositional sense when they are complements of verbs.

(12) Trovo quell'imbrogliona di Marta.
only 'I find that cheater Marta.'
not 'I find Marta a cheater.'
(cf. Trovo Marta un'imbrogliona.
'I find Marta a cheater.'

Fifth, they behave like R-expressions with respect to the Binding Conditions. Thus they can bind an anaphor or a pronoun, but they must themselves be free.

(13) a. Quel matto di Giorgio parla fra sè e sè.
'That madman George talks to himself.'

b. [Quel matto di Giorgio]$_i$ vuole che io lo$_i$ inviti.
'That madman George wants me to invite him.'

c. *Gli$_i$ ho detto che [quel matto di Giorgio]$_i$ non poteva venire.
'I told him that that madman George couldn't come.'

I conclude that phrases like **quel matto di Giorgio** are NPs.

2. Internal Structure

The internal structure of these phrases is the following:

(14) $_{N''}$[quel $_{N'}$[$_{N}$[matto] $_{pp}$[di $_{N''}$[Giorgio]]]]

This is precisely the structure we would expect with X Bar Theory.

The evidence that **di Giorgio** in (14) is a PP sister to the head N is scarse. Like other PP complements to N, this PP cannot extrapose.

(15) *Quel ficcanaso non si vede piu di Stefano.
'That busybody isn't seen anymore of Stephen.'
(cf. *Quei libri non si vedono piu di quello scrittore.
'Those books aren't seen anymore by that writer.')

Like other instances of the P **di**, we find complementary distribution with the corresponding inflected P (see Napoli and Nevis 1986 for an analysis of inflected Ps in Italian). In (16) **dei** is the inflected P.

(16) Quei cretini dei ragazzi non ci sono.
'Those cretin boys aren't here.'

Like other Ps, the **di** governs oblique Case. To see this we must use a phrase containing a pronoun after the **di** since Italian, like English, does not show audibly distinguishable Case on nonpronominal NPs. However, not all Italians find these phrases perfectly acceptable with a pronoun following the **di**. Still, all agree that if a pronoun is present, it must be oblique and never nominative.

(17) Quel cretino di te/*tu!
obl/nom
'That cretin you!'

The **di** phrase here, however, does not have the syntactic flexibility of many other **di** phrases. I list some characteristics that our **di** phrase lacks. None of these characteristics are defining characteristics for **di** PPs, however, since none of these characteristics are common to all **di** PPs, even excluding from consideration the **di** phrase at issue.

It cannot be questioned.[1]

(18) *Di chi è quel furbo?
'Of whom is that sneak?'
(cf. Di chi è quel libro?
'By whom is that book?')

It cannot correspond to a clitic **ne**.

(19) *Ne ho visto quel matto.
'I saw that madman him.'
(cf. Ne ho visto la foto.
'I saw that photograph of him.')

It cannot appear across the copula from the N and its specifier.[2]

(20) *Quello stupido è di Franco.
'That stupid person is (of) Frank.'
(cf. Quel libro è di Franco.
'That book is by Franco.')

It cannot prepose.

(21 *Di Giorgio non si vede più quel matto.
'George one doesn't see anymore that madman.'
(cf. Di quello scrittore non si vedono più i libri.
'By that writer one doesn't see books anymore.')

In all four of these properties, our phrase is similar to NPs such as **la città di Sassari** 'the city of Sassari.'

(22) *Di cosa è la città?
'Of what is the city?'

*Ne ho visto la città.
'I saw the city of it.'

*La città è di Sassari.
'The city is of Sassari.'

*Di Sassari mi piace la città.
'Of Sassari I like the city.'

It is significant that the **di** phrase in **quel matto di Giorgio** and the di phrase in **la città di Sassari** are syntactically inflexible in the same ways since I argue in Napoli 1987 that NPs like **la città di Sassari** also involve predication and I attribute the syntactic inflexibility to this fact.[3] Thus the data in (18) - (21) do not offer evidence against the analysis in (14).

The most likely alternative to a PP analysis of the **di** phrase in these NPs is one in which the **di** is a Case marker and not a true preposition (see Stowell 1981 for one side of this argument with respect to English **of**, but others, such as Zubizarreta 1985, for the other side). This alternative is interesting to this paper only insofar as one could argue with the Case marker analysis of **di** that the external analysis of the **di** phrase is NP, instead of PP, since in that instance the NP which is the maximal projection headed by **Giorgio** would be **di Giorgio** and, thus, the maximal projection of **Giorgio** would, in fact, c-command the head N **matto**. The Case marker analysis is not viable, however. As we noted above with respect to (16), the **di** of our phrases appears in complementary distribution with the corresponding inflected P, just as the true preposition **di** does. Now since the inflected P is a true preposition and not a Case marked article (see Napoli and Nevis 1986 for arguments to this effect), we need to admit the structure in (14) as the proper analysis of sentences like (16), with the inflected P. But given that a PP can follow the head N of our NPs in some instances (like (16)), we gain nothing from a theoretical point of view by arguing that **di** in other examples is a Case marker and not a true P. Furthermore, since Italian has a true preposition homophonous with the **di** of these phrases, in the spirit of avoiding proliferation of homophonous items that have similar distribution, I choose to analyze the **di** here as a preposition. Throughout this paper I point out all places where the choice between the analysis of **di** as a Case marker and **di** as a preposition is relevant, for the benefit of the reader who wishes to pursue the Case marker analysis.

Evidence that the material between the initial specifier and the **di** phrase in a structure such as (14) is an N is offered by four facts. First, the types of specifiers that occur before this material are those that Ns typically take.[4] Second, a PP complement (which is what the **di** phrase in (14) is) is a typical type of N complement. Third, the lexical items that can appear in this slot are only those that can appear as the head of regular NPs (see also Milner 1978, p. 175). Thus, adjectives that can appear in regular NPs in the absence of an N only with an elliptical sence cannot appear in this slot, nor can Vs or Ps.

(23) Preferiso il verde.
'I prefer the green (one/color/thing/...).'

(24) *una verde di frutta 'a green of a fruit'

Fourth, the material in this slot can be filled by more than one word, in which case we alsways get an N plus a complement that is the typical type of complement an N takes. In (25) we have the AP **ripulito.** AP appears as a complement only to N unless it is used as a predication.

(25) quel cafone ripulito di Giorgio
'that polished up bumpkin George'

I conclude that the material between the initial specifier and the **di** phrase in (14) is an N.

Other empirically based arguments for the structure in (14) are hard to come by. However, there is a good theoretical reason for positing 14: there is simply no other reasonable structure available. Notice that if the first N of the construction was not the syntactic head, we would have no way of assigning Case to it. That is, if the first N were the head of an NP inside the overall NP, as in (26), there would be no Case assigner for the contained NP.

(26) $_{N''}$[$_{N''}$[quel matto] $_{PP}$[di Giorgio]]]

In (26), the highest N" would protect the NP **quel matto** from outside Case assignment, since maximal projections are barriers to government and Case can be assigned only by a governor.[5] Furthermore, if **quel matto** were a maximal projection, then **Giorgio** would have to be the head of the whole NP. We now have a question as to what **di** is doing here. If **di** is a preposition, it surely should not be introducing a nonmaximal projection of NP (but see Napoli and Nevis 1986). On the other hand, one could argue that the obvious function of the **di** in our phrases is to assign Case to the NP which follows it (see the discussion following (22) in the text above). But if the material following **di** in these phrases were not an NP, but simply an N which was the head of the overall NP, there would be no need for the **di**, since the Case governor for the overall NP would assign Case to this N by percolation. Alternatively, if the material following the **di** were not the head N but a full NP, the construction in (26) would be headless, contrary to X' theory (but see Napoli 1985 and Williams 1974, p. 139). In fact, the material following the **di** is a full-fledged NP, which can have its own specifier and other complements.

(27) quella carogna del tuo dottore
'that scoundrel of your doctor'

(26) cannot be correct.

Another alternative I raise simply because it comes up in the literature is that in (28).

(28) $_{N''}[\ _{N''}[$quel matto di$]\ _{N'}[\ _{N}[$Giorgio$]]]$

(This is the structure of the pseudopartitive in Selkirk 1977, who labels the top node N"' instead of N". This is also one of the two possible structures proposed in Janda 1980 for the NPs studied there - see note 3 above.) Once more this structure fares no better with respect to how to assign case to the N" **quel matto di** and as to why we find a **di** at all. Furthermore, (27) shows that **Giorgio** cannot have N' as its highest analysis, but must have N" as its highest analysis, since we can get a specifier on the material following **di**. Also, the structure in (28) is completely inadequate for handling examples like (16), repeated here as (29), in which we find the inflected preposition.

(29) Quei cretini dei ragazzi non ci sono.
'Those cretin boys aren't here.'

There is still another argument against (28), based on the fact that (28) analyzes the nominal that follows the **di** as the head N of the construction. We expect with this analysis that this N will manifest the Case that is assigned to the overall NP. However, as mentioned above with respect to

(17), we never get nominative Case even if the overall NP is in subject position of a tensed S, the position typically assigned nominative.[6]

(30) Quel cretino di te/*tu ha rovinato tutto di nuovo.
obl/nom
'That cretin you ruined everything again.'

(28) cannot be correct.

I conclude that (14) is the correct analysis of these NPs. Given (14) we can see that **Giorgio** neither c-commands nor is external to the maximal projection of **matto** in the phrase **quel matto di Giorgio.**[7]

3. Semantics

The semantic properties of these NPs are complex. I will first discuss those properties in common among all NPs of this type. Then I will discuss properties that only some NPs of this type have.

3.1 Semantic Properties in Common

The head N of the overal NP in (14) acts as a predicate to the NP introduced by **di**, which acts as its subject argument, and the initial specifier, as expected, specifies the entire NP.

The claim about the specifier is simpler to justify, so I will begin with it. When the specifier of these NPs is definite, the entire NP is interpreted as definite. In (31) - (32) the subject of the subordinate clause must be indefinite when it is understood as coreferential with the subject of the matrix clause. The initial specifier satisfies this requirement, whether the NP following **di** is definite, as in (31), or indefinite, as in (32).

(31) Per quanto quel cretino di Giorgio sembri onesto,
non lo è.
'Honest though that cretin George may seem, he
isn't.'

(32) Per quanto quella carogna di dottore sembri intelligente, non lo è.
'Intelligent though that scoundrel of a doctor may seem, he isn't.'

Likewise, when the initial specifier of these NPs is indefinite, the entire NP is indefinite. In (33)-(34), if the position following **Antonio** is filled by an NP, it must be a predicative NP. In general indefinite NPs are potential predicates whereas some uses of definite NPs are not, with context being the relevant factor for potential interpretation of definite NPs as predicates.[8] The initial specifier on our NPs satisfies this requirement. Thus (33), with **una**, the indefinite article, is good, whereas (34), with **quella**, the definite demonstrative adjective, is not.

(33) Considero Antonio una bellezza di marito.
'I consider Antonio a beauty of a husband.'

(34) *Considero Antonio quella bellezza di marito.
'I consider Antonio that beauty of a husband.'

Notice that in both (33) and (34) the NP following **di** is indefinite. Thus the indefiniteness of the overall NP is due to the initial specifier, not to the specifier of the NP following di.

I will now argue that the head N of our NPs, as analyzed in (14), functions as a predicate taking the NP following **di** as its subject.

My first three arguments all have the same form: I show that the N I am calling a predicate and the NP I am calling its subject exhibit properties that typically occur only between predicates and subjects.

First, as Alinei 1971 has pointed out, there must be a matching of selectional restrictions between the N predicate and the NP subject which is identical to the matching of selectional restrictions we would find in the corresponding sentences with the copula.

(35) *quella lagna di monumento
'that complainer of a monument'
(cf. quella lagna di tua sorella
'that complainer of your sister')

(36) *quel polpettone di motore
'that confused mess of a motor'
(cf. quel polpettone di articolo
'that confused mess of an article')

(37) *quella bruttura di organizzazione
'that filth of an organization'
(cf. quella bruttura di mobile
'that filth of a piece of furniture')

Second, if the predicate is one that may vary for gender and/or number, it must agree with the NP for those features, just as it would in a copula construction, as Milner 1978 (p. 192) has pointed out for French.

(38) quello stronzo di Mario
m.s. m.s.
'that shit Mario'

(39) quella stronza di Giuliana
f.s. f.s.
'that shit Giuliana'

(40) quel coglione di suo fratello
m.s. m.s.
'that jerk of his brother'

(41) quei coglioni dei suoi fratelli
m.p. m.p.
'those jerks of his brothers'

Third, the N is understood to assign a property to the NP in the same way that predicates assign properties to their subjects in more familiar predicational structures (such as clauses). Milner 1978 says that the NP "apparaît en quelque sorte comme le sujet logique de l'ensemble" (p. 174-5). Thus in (42a) we understand that the dress in question is shabby and in (42b) we understand that the doorman in question is rude.

(42) a. questo straccio di vestito
'this rag of a dress'

b. quello sgarbato del portiere
'that rude person doorman'

If the NP following **di** were an argument of the nead N but not a subject argument (as has been argued, for example, for the ergative verb constructions in Burzio 1981), we would have no explanation for why the same theta role is assigned to the NP following **di** as that assigned to the external argument of these same predicates in copular constructions or elsewhere.

(43) a. quel fesso di Paolo
'that dumbo Paolo'

b. Paolo è un fesso.
'Paolo is a dumbo.'

c. Considero Paolo un fesso.
'I consider Paolo a dumbo.'

Fesso assigns the same theta role to **Paolo** in all three examples of (43). If we claim that **Paolo** is the subject argument of **fesso** in (43b) and (43c) but not in (43a), we give up correlating theta roles with particular arguments of given predicates in D.S. Furthermore, if **fesso** is a predicate in (43a) and if all predicates must have subjects (as argued in Rothstein 1983), then **Paolo** must be the subject of **fesso** by default.

A fourth argument for the predicate status of the head N is that only Ns which can function as predicates are accepted in this position (see also Milner 1978, p. 175). While almost any N can be interpreted as a predicate given an appropriate context (see note 8 above), proper names and pronouns are extremely resistent to being interpreted as predicates. As expected, proper nouns and pronouns cannot occur in head N position.

(44) *quello Carlo d'uomo
'that Carlo of a man'

(45) *un tu/te di persona
nom/acc
'a you of a person'

One might object to 44 and 45 being taken as evidence for the predicate status of the head N of our NPs by noting that 44-45 could be out for the extraneous reason that proper names and pronouns do not take specifiers. There are other classes of nouns, however, that can easily occur with specifiers but that do not lend themselves to being interpreted as predicates in these NPs. The head N of these NPs is interpreted as an evaluative judgment of the NP following **di** (see section 3.2 below). The names of inanimate objects which are typically used for specific purposes are not generally employed as predicates with the sense of an evaluative judgment. Therefore, these classes of nouns are also barred from head N position in our NPs.

(46) *una chiave di attrezzo
'a wrench of a tool'

*una lampada di mobile
'a lamp of a piece of furniture'

*un quadro d'opera d'arte
'a picture of a work of art'

As expected, those inanimate objects which are colloquially used with the sense of an evaluative judgment can appear in head N position in our NPs.

(47) un fiore di moglie
'a flower of a wife'

una rosa di moglie
'a rose of a wife'
(cf. *un tulipano di moglie
'a tulip of a wife')

The same selection of head N occurs in the English counterpart to the NP in 47 (see also Napoli 1987).

(48) a pistol of a lecturer
(cf. *a revolver of a lecturer)

On the other hand, Ns which easily lend themselves to being interpreted as predicates of the appropriate type easily appear as the head N in our NPs. Any of the good examples in this paper can attest to this point.

The next two arguments that our NP consists of a predicate and its subject have the same form: I show that with the analysis here expectations about reference and agreement phenomena arise and, indeed, are realized.

First, if the NP following di is the subject of the head N we should expect the referent of the overall NP to be more closely related to the meaning of the NP following **di** than to the meaning of the head N. That is, the meaning of the overall NP is the meaning of the NP following **di** with the property assigned to it by the predicate head N. This is the case. In (49) we understand the subject of the sentence to want to marry a special kind of girl (one who is as fine as a flower) rather than to want to marry a special kind of flower.

(49) Vorrei sposare un fiore di ragazza.
'I'd like to marry a flower of a girl.'

Even more convincing examples occur in the comparable English construction. In (50) we certainly don't mean that the subject of the sentence says anything about hell.

(50) He tells one hell of a good story.

(49) and (50) are telling, as well, with respect to the theta role of the whole NP. We see that the NP following **di** must meet the selectional restrictions that the larger context imposes on the whole NP, as Milner 1978 points out for French. Thus it is the NP following **di** which can be said to bear the theta role of the whole NP. In (49), for example, **ragazza** bears the theta role of theme with respect to **sposare**.

Second, given the meaning of these NPs we might expect speakers to treat them as though the N of the NP following di is the real syntactic head in grammatical processes that involve semantic features of the syntactic head. For example, agreement processes in Italian are sensitive to the features of person, number and gender - all features which correlate closely to semantic properties of the referent of the NPs undergoing agreement. Let us look, then, at agreement processes.

Subject-Verb Agreement in Italian makes agreement for person between a tensed verb and its syntactic subject (that is, [NP,S]). In Italian morphological person on the verb is correlated with semantic person of the referent of the verb's syntactic subject except in formal speech (where a morphological third person correlates to a semantic second person) and in the indefinite subject si construction (where a morphological third person can correlate with various semantic persons, see Napoli 1973). However, Subject-Verb Agreement takes place between all syntactic subjects and their tensed verbs, whether the syntactic subject is an argument of that verb of not (witness passives involving idiom chunks, and, of course, dummy subjects in languages like French and English). Thus Subject-Verb Agreement is not based on a semantic relationship between the syntactic subject and its VP, but is merely a mechanical process involving features (person and number) which irrelevantly (with respect to Subject-Verb Agreement) have semantic correlates. Therefore, I predict that Subject-Verb Agreement will take person strictly off the syntactic head N of our NPs. To test this we must find an NP of the type in (14) in which the two Ns have different person, and see whether this overall NP as the syntactic subject of a verb calls for person agreement with the head N or with the NP that follows the di. Since NPs in Italian are all third person except for pronouns, we must use a pronoun either as the head N or as the NP following di in order to allow for a contrast in person. However, pronouns cannot appear in head N position, as explained above with regard to (45). And, as we saw above with (17), pronouns are rejected by many Italians in the position following di in these NPs. Thus, unfortunately, we must test with sentences which are not acceptable to all speakers. Still, all speakers agree that the morphological marking on the verb in these cases must be third person. We can use (30), repeated here as (51), to see this.

(51) Quel cretino di te ha/*hai rovinato tutto di nuovo.
3 2 3 2
'That cretin you ruined everything again.'

Subject-Verb Agreement, as predicted, picks out the syntactic head of our NPs, regardless of a contrast in person between head N and the NP following di.

There is another feature that figures in Subject-Verb Agreement in Italian: number. The analysis here predicts that if we find NPs in which the head N and the NP following di have different number, number agreement will be strictly with the syntactic head N. This prediction holds, as John Swales (personal communication) has pointed out to me. Thus while an NP such as those in (52a) has two readings, on both readings Subject-Verb Agreement is with the syntactic head, as in (52b).

(52) a. i cretini della Mafia
'the cretins of the Mafia'
nonpredicational reading: the cretins who are
(restrictive) part of the Mafia
predicational reading: the Mafiosi, all of whom
(nonrestrictive) are cretins

quei fessi della Chiesa
'those dumbos of the Church'
nonpredicational reading: those dumbos who are
(restrictive) clerics
predicational reading: those clerics, all of
(nonrestrictive) whom are dumbos.

b. I cretini della Mafia non capiscono/*capisce
pl. sg. pl. sg.
l'onore.
'Those cretin Mafiosi don't understand honor.'

Quei fessi della Chiesa oppongono/*oppone ancora
pl. sg. pl. sg.
il divorzio.
'Those dumbo clerics still oppose divorce.'

On the whole, then, Subject-Verb Agreement treats our NPs as we would expect it to.

There is yet another feature for which agreement can be made, and that is gender. This feature does not show up on a tensed verb, but rather on certain participles and almost all adjectives. Certainly with respect to adjectives, the fact that they agree with an N or NP is related to the fact that they modify or, in the case of predicate adjectives, predicate, the N or NP. Thus this agreement process is truly semantically based, unlike Subject-Verb Agreement. We predict, therefore, that speakers might make gender agreement with the NP following di rather than with the head N in our NPs when it is modified or predicated by some adjective outside the overall NP. If we could find an NP of the type in (14) in which the head N and the NP following di contrasted in gender, we could test whether an element outside the NP which is to agree in gender with that NP takes the head N or the NP following di to agree with. We can, in fact, find examples where the head N and the NP following di contrast in gender.

(53) quell'angelo di tua moglie
m. f.
'that angel of your wife'

una peste di bambino
f. m.
'a wretch of a boy'

When such NPs are used in sentences where gender agreement with some adjective or participle outside the NP is called for, speakers are not happy. They typically hesitate in giving a judgment on a prepared sentence or in producing the required sentence on their own. Then more often than not they choose agreement with the NP following di. Milner (1978, p. 193) likewise reports for French that agreement is with the NP following de in these instances.

(54) a. Un gioiello di moglie non sarebbe incinta/??incinto continuamente.
'A jewel (m.) of a wife (f.) wouldn't be pregnant (f./??m.) continually.'

b. Questo straccio di gonna è troppo ?lungo/lunga.
'This rag (m.) of a skirt (f.) is too long (?m./f.).'

c. Una bestia di avvocato è arrivato/??arrivata.
'A fool (f.) of a lawyer (m.) has arrived (m./??f.).'

d. Quel diavolo di Maria è ?arrivato/?arrivata.
'That devil (m.) of Maria (f.) has arrived (?m./?f.).'

One might object that the oddity of **incinto** 'pregnant' in the masculine adds to the problem in (54a). Such an objection does not hold, however, since sentences like (55) are perfectly acceptable.

(55) Il donnone è incinto per la quinta volta.
'The big woman is pregnant for the fifth time.'

The problem in (54a-d), then, is due to the contrast between genders of the two Ns within the NP, where the overall NP's referent is determined more by the referent of the NP that follows **di** than by the head N. While not all speakers may give precisely the acceptability judgments I have indicated in (54a-d), no speakers feel completely at ease with whatever choice of agreement they make in these examples. And all speakers say they'd avoid using such sentences. The confusion speakers feel at sentences like (54a-d) is expected, given the analysis here.

3.2 Other Semantic Properties

First, when the NP following **di** within the overall NP is indefinite, that NP is understood as having a referent picked out by the fact that it bears the property which it itself describes by definition (that is, by the sense of the N) and it is in the capacity of having this property that it is assigned the additional property of the sense of the head N. The overall NP, therefore, can be used itself as a predicate assigning a property to some argument, as well as being used referentially.

(56) Maria è un fiore di moglie.
'Maria is a flower of a wife.'

Un fiore di moglie è una meraviglia.
'A flower of a wife is a marvel.'

In either use in (56), it is in the property of being a wife that the NP following **di** is assigned the added property of being flowerlike. The NP following **di**, then, typically does not have a specifier of its own, just as it typically would not in the corresponding sentence with a copula (see Gunnarson 1986, p. 35, for French). Consider (57), suggested to me by Giulio Lepschy (personal communication).

(57) a. quell'ignorante di dottore
'that ignoramous of a doctor'

b. quell'ignorante del dottore
'that ignoramous doctor'

The referent of (57a) is that person who in the property of being a doctor is ignorant. In other words, he doesn't know his job, but perhaps as a musician he's a cultured person. (57a) contrasts with (57b). Here the NP following the inflected P del is definite and the whole NP is understood as having as its referent someone who is a doctor but the property of being a doctor is accidental to the assignment of the additional property of being an ignoramus. It may well be that he's a very fine doctor, but he's culturally a jerk. In (57a), then, the property of being a doctor is crucial to the assignment of the property of being an ignoramus. But in (57b) the property of being a doctor is accidental to the overall NP and is merely one way of identifying the person who happens to be an ignoramus.[9]

We expect, then, that if the sense of the indefinite NP following di is not an appropriate property to be assigned the additional property of the predicate N preceding di, the overall NP will be rejected. This is the case, as poined out to me by Peter Hook (personal communication).

(58) a. *un fesso di neonato
'a dumbo of a newborn'

?un fesso di ragazzino
'a dumbo of a little boy'

un fesso d'uomo
'a dumbo of a man'

b. ??quel fesso del neonato
'that dumbo newborn'

c. ?*un donnaiuolo di dottore
'a womanchaser of a doctor'

d. il donnaiuolo del dottore
'the womanchaser doctor'

In (58a) we see that in someone's property of being a newborn it is absurd to assign the property of being a dumbo. That is, we don't typically characterize the behavior of newborns as being like that of dumbos or not. But the absurdity decreases as the person's age and ability to be a dumbo at that stage of life increases. Thus characterizing little boys as being dumbos is not so silly and characterizing adults as being dumbos is utterly normal. The NP following di in (58b), however, is definite, and while (58b) is certainly odd, it is not absurd in the same way as the first sentence of (58a). With the definite following di we are picking out that newborn that happens to be a dumbo. The oddness is in the judgment: how on earth might someone come to the conclusion that a given newborn is a dumbo? But the possibility arises that this newborn may have done something which is stupid, even for a newborn. Such a situation is odd but not definitionally absurd, in contrast to the idea of setting up classifications for newborns into dumbos and nondumbos.

In (58c), with an indefinite following di, we see that the property of being a doctor is most probably unrelated to the property of being a womanchaser. That is, it is hard for us to imagine a context in which in the capacity of a doctor a certain person is a womanchaser. But (58d) shows us that a doctor can happen to be a womanchaser, just probably not as a result of his capacity as a doctor.

We are now in a position to explain the following array of data. Our NPs can be definite or indefinite overall and the NP following di can be definite or indefinite (see the discussion of (31)-(34)). There are four logically possible combinations of definiteness on the specifiers in these NPs, then. But only three of them occur.

(59) quell'egoista del direttore
'that egotist director'

(60) quell'egoista di direttore
'that egotist of a director'

(61) *un'egoista del direttore
'an egotist director'

(62) un'egoista di direttore
'an egotist of a director'

Given that when the NP following **di** is a definite NP it is understood as a referential NP, and given that the referent of the overall NP is determined more by the referent of the NP following **di** than by the head N, we can see that if the NP following **di** is definite, the entire NP must be definite or we will be assigning contradictory features to the referent of the overall NP. Thus (61) is unacceptable because we know we are talking about the referent of **del dottore** as though it is anaphoric to the discourse but at the same time the indefinite specifier **un'** tells us we are talking about the referent of the overall NP as though it is new to the discourse.[10] Thus when the NP following **di** is definite, the entire NP must be definite.[11]

When the NP following **di** is indefinite, however, we have a different type of situation because the NP following **di** now functions to pick out a referent in its capacity of having a certain property rather than as just happening to be identifiable by that property. Since it is a property assignment we are dealing with, there is no information given us by this NP as to its anaphoricity or lack of anaphoricity within the discourse. And, as noted above, we typically have no specifier on the NP following **di** at all. Thus the specifier of the overall NP is the only specifier, or, at least, the only specifier which indicates the feature of discourse anaphoricity. Therefore, if the overall NP is new to the discourse, an indefinite specifier in initial position is appropriate and if the overall NP is anaphoric to the discourse, a definite specifier in initial position is appropriate. No potential for contradiction internal to the overall NP arises.

Another semantic property of our NPs is that the head N gives an evaluative judgment of the NP following **di**. Typically this judgment is negative, but it can be positive (see Alinei 1971). If the overall NP is definite, the most likely reading is of negative judgment, so that (63), for example, out of context would favor an ironic reading (see Lepschy and Lepschy 1977).

(63) quel tesoro di Giorgio
'that treasure George'

If the overall NP is indefinite, the tendency for a negative judgment disappears, so that (64)-(65) have no ironic flavor.

(64) un gigante d'uomo
'a giant of a man'

(65) un gioiello d'amico
'a jewel of a friend'

I have no explanation for these facts.

Another semantic property of these NPs is that some speakers can have a complement on the head N which semantically belongs to the NP following **di.** All such speakers that I found are also speakers of English, and since English does this rampantly (see Hall 1973), I do not know if there is interference here. Milner 1978 gives many examples which display the same phenomenon in the comparable French construction.

(66) il tuo cretino di fratello
'your cretin of a brother'

(67) your asshole of a sister
French: mon crétin de mari
'my cretin of a husband'

Notice that the phenomenon in (66)-(67) may be taken as one more indication of the strength of the NP following **di/of** with respect to the determination of the referent of the overall NP. (See the discussion around examples (49)-(56) above.)

4. Conclusion

We have seen a particular structure, NPs like **quel matto di Giorgio** in Italian, in which the subject argument of a predicate neither c-commands the predicate nor is external to the maximal projection of that predicate. I conclude that the c-command and externality requirements on predication found in the linguistic literature are wrong.

Many questions arise now. First, one might wonder what led linguists to propose the c-command and externality requirements on the subject argument of a predicate. In Napoli 1987 I argue that the crucial evidence comes from the fact that many objects of prepositions are not available as subject arguments to predicates outside their PP. I show there that the relevant restriction is that in most instances of secondary predication the subject argument of the secondary predicate must receive a theta role from some source other than the secondary predicate in addition to the theta role assigned by the secondary predicate.[12] With that restriction in mind, we can see that those objects of prepositions which do not receive theta roles independently of the secondary predicate cannot serve as the subject argument to that secondary predicate. But those objects of prepositions which receive theta roles independently of the secondary predicate can serve as the subject argument to that secondary predicate.

Second, if one were to follow the lead of Safir 1985 and propose that there is only one kind of indexing in grammar, then one would want to have predication indexing and binding indexing be sensitive to the same constraints. There is evidence that, in fact, predication and binding present the same sorts of questions for indexing theory (see Zubizarreta 1985 and Napoli 1987).

Third, the relevance of the notion of c-command and of the external/internal distinction to grammar is now open to question. I suggest in Napoli 1987 that c-command is a necessary notion in grammar, but that its applications are much more restricted than originally proposed in early GB literature (such as Chomsky 1981). Furthermore, I argue that the external argument notion is not the useful notion, but, rather, should be replaced in GB by the notion of subject argument (contra Williams 1984b), a notion which cannot be identified by any structural position.

These results, if correct, are perhaps not welcome but at least not unexpected. Predication is a semantic phenomenon involving the semantic relationship of property assignment. That predication should allow a wide range of syntactic relationships between a predicate and its subject argument may be taken a symptomatic of the proper division between semantic phenomena and syntactic phenomena.

NOTES

* I thank Josh Ard, Ivonne Bordelois, Marcel Danesi, Dick Hudson, Giulio Lepschy, Marina Nespor, and John Swales for comments on early versions of this work. Deep gratitude goes to Peter Hook and Barry Miller for prolonged conversations about the issues raised here.

[1] I exclude from consideration here the irrelevant and grammatical reading of (18) in which the **di** phrase is a possessive, as in (i), suggested to me by Marcel Danesi.

(i) Guarda quanto è furbo quel bambino. Di chi è quel furbo?
'Look how sneaky that child is. Whose is that child?'

[2] Again I exclude from consideration the irrelevant and grammatical reading of (20) in which the **di** phrase is a possessive, parallel to the situation in note 1 above.

[3] The English counterparts to both of these types of Italian NPs are also syntactically inflexible, as noted in Janda 1980, p. 332.

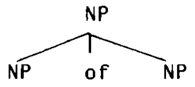
(i) *Who did Mary invite that fool of t?
(cf. Mary invited that fool of a doctor.
but: Who did Mary invite the sister of t?)

(ii) *What did Mary see the city of t?
(cf. Mary saw the city of Taormina.
but: What did Mary see the destruction of t?)

Napoli 1987 argues that the English NPs in (i) and (ii) also involve internal predication. In these examples it is the NP that follows **di/of** that bears the theta role assigned to the whole NP from its outside context (see the discussion of (49)-(50) below in the text). Thus the NP **that fool of t** in (i) is not L-marked (in the sense of Chomsky 1986). Therefore the NP node dominating **that fool of t** is a barrier for the t. Therefore we have a violation of subjacency in (i), and similarly in (ii) and in (18), (19), (21), and (22) in the text.

Janda 1980, on the other hand, takes the syntactic inflexibility of examples like (i) and (ii) as evidence that the **of** and the NP that follows it do not form a PP. He examines NPs like **a friend of John's** and argues that they have an analysis either like that in (28) below in the text or like that in (iii).

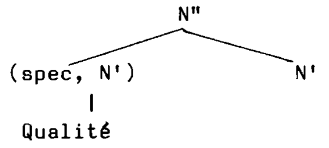
(iii)



He explains the inability to move of the NP following **of** by appealing to the A/A principle, since this NP would be immediately dominated by another NP node in either analysis. Since he compares his NPs to NPs like those in (i) and (ii), I assume he would propose the same two possible analysis for these NPs. The real issue here with respect to the position of **of** (as in (14) or (28) or Janda's (iii)), it seems to me, is whether **of** is a P in (i) and (ii) or a Case-marker. For English I take no stand on this issue, since neither analysis of **of** crucially affects my claim that a subject argument need not c-command its predicate, given that **of** isn't present in some of the types of predication structures that I use to make my point in Napoli 1987. And notice that even in Janda's two possible analyses, the subject argument would not be external to the maximal projection of its predicate. For Italian I take the stand that **di** is, in fact, a preposition, for reasons given immediately below in the text.

Milner 1978 also observes the syntactic inflexibility of the comparable French construction and partly because of it he argues that in N_1 **de** N_2 the **de** N_2 is not a PP and that the N_2 is an N' and not an N". The structure he offers is given in (iv) (see his p. 190), where N_1 is under the specifier slot and N_2 is under N'.

(iv)



For Italian it is clear that N_2 is an N" not an N'; see (27) below in the text and many other such examples. However, Milner gives no French examples comparable to (27). It may be that Italian and French differ on the crucial data here. (See also Milner's ungrammatical example 6.23 on p. 235, whose Italian counterpart is perfectly grammatical.)

[4] Marta Lujàn and Osvaldo Jaeggli (personal communication) told me that in Spanish **muy** 'very' can appear with the head N in these NPs. In general **muy** and its Italian counterpart, **molto**, do not appear with Ns, but with As. However, **molto** may appear with an N which is used predicatively, as can **muy** in Spanish.

(i) E molto uomo.
'He's very much a man.'

Thus I do not take the ability of **muy** to appear in these NPs in Spanish as evidence against the claim that the head of the whole phrase is an N.

[5] Italian, unlike English, does not have a rule assigning genitive Case to NPs in specifier position of NP.

[6] Again, as with (17), some speakers reject (30) since they reject all pronouns from these constructions. However, as with (17), all say **tu** is much more awful that **te**.

[7] If one argued that the **di** of these phrases is a Case-marker and if Case-markers are not Ps but inflections on NP (so that in (14) we have an NP sister to the head NP rather than a PP sister), then **Giorgio** would c-command **matto**. I take **di** to be a P rather than a Case-marker for reasons already stated in the text. Therefore, I accept (14) here. Let me point out, however, that if one could defend the proposal that the PP in (14) is actually an NP, this analysis would still be consistent with the theory of predication in Napoli 1987.

[8] Rothstein (1983, p. 104) says that when an NP contains evaluative material it is more likely to be open to the interpretation as a predicate. See also Hawkins 1978, Williams 1980, Woisetschlaeger 1983, and Safir (1985, p. 170).

[9] The remarks on the interpretation of (57b) hold for Italian only. For a discussion of the semantics of the complex nominal used in the English gloss of (57b), see Napoli 1987.

[10] See Rando and Napoli 1978 for a discussion of this notion of anaphoricity.

[11] This is, as far as I know, a new explanation for the matching of specifiers, traditionally accounted for with Bembo's rule of corresponding articles.

[12] This claim is close to being in complete opposition to that of Bresnan 1982. The differences are discussed and evaluated in Napoli 1987.

REFERENCES

Alinei, M. 1971. "Il tipo sintagmatico **quel matto di Giorgio**, in **Grammatica trasformazionale italiana**, ed. by Mario Medici and Raffaele Simone, Rome: Bulzoni. 1-12.

Bresnan, J. 1982. "Control and complementation", **Linguistic Inquiry** 13, 343-434.

Burzio, L. 1981. **Intransitive verbs and Italian auxiliaries**, Ph.D. dissertation. MIT.

Chomsky, N. 1981. **Lectures on Government and Binding**, Dordrecht: Foris Publications.

Chomsky, N. 1986. **Barriers**, Cambridge, Mass: MIT Press.

Gunnarson, K.-A. 1986. "Predicative structures and projections of lexical dependence", **Linguistic Inquiry** 17, 13-48.

Hall, R.A. Jr. 1973. "The transferred epithet in P.G. Wodehouse", **Linguistic Inquiry** 4, 92-4.

Hawkins, J. 1978. **Definiteness and indefiniteness: a study in reference and grammaticality prediction.** Atlantic Highlands, N.J.: Humanities Press.

Janda, R. 1980. "On a certain construction of English's", in Bruce Caron et al. (eds.) **Proceedings of the sixth annual meeting of the Berkeley Linguistics Society**, Berkeley: BLS. 324-36.

Lepschy, A.L. and G. Lepschy. 1977. **The Italian language today**. London: Hutchinson & Co.

Milner, J.-C. 1978. **De la syntaxe a l'interprétation, quantités, insultes, exclamations**, Paris: Editions du Seuil.

Napoli, D.J. 1973. **The two si's of Italian: an analysis of reflexive, inchoative, and indefinite subject sentences in modern standard Italian.** Ph.D. dissertation. Harvard. Distributed by IULC in 1976.

Napoli, D.J. 1985. "Verb phrase deletion in English: a base generated analysis", **Journal of Linguistics**, 21, 281-319.

Napoli, D.J. 1987. **Predication Theory: A case study for Indexing Theory**, Cambridge: Cambridge University Press. to appear.

Napoli, D.J. and J. Nevis. 1986. "Inflected prepositions in Italian", to appear in Arnold Zwicky and Ellen Kaisse (eds.) **Phonology Yearbook**, vol. 4.

Rando, E. and D.J. Napoli. 1978. "**There** sentences in English", **Language** 54, 300-13.

Rothstein, S. 1983. **The syntactic forms of predication**, Ph.D. dissertation, MIT, distributed by IULC in 1985.

Safir, K. 1985. **Syntactic chains**, Cambridge: Cambridge University Press.

Schein, B. 1982. "Small clauses and predication", unpublished manuscript, MIT.

Selkirk, E. 1977. "Some remarks on noun phrase structure", in **Formal Syntax**, ed. by Peter Culicover et al., New York: Academic Press, Inc., 285-316.

Stowell, T. 1981. **Origins of phrase structure**, Ph.D. dissertation, MIT.

Williams, E. 1980. "Predication", **Linguistic Inquiry** 11, 203-38.

Williams, E. 1984. "There-Insertion", **Linguistic Inquiry** 15, 131-53.

Williams, E. 1984b. "Grammatical Relations", **Linguistic Inquiry** 15, 639-73.

Woisetschlaeger, E. 1983. "On the question of definiteness in 'An old man's book'", **Linguistic Inquiry** 14, 137-54.

Zubizarreta, M.-L. 1985. "The relation between morphophonology and morphosyntax: the case of Romance causatives", **Linguistic Inquiry** 16, 247-290.

Rafael A. Nuñez Cedeño

Structure-preserving Properties of an Epenthetic Rule in Spanish

A quick look at the literature of the Hispanic Caribbean, or even at the linguistic literatures of Spanish America, shows that the aspiration or deletion of syllable-final /s/ is the topic most thoroughly researched by phonologists of different theoretical inclinations. Numerous descriptions have been provided attempting to explain the phenomenon. Among these, perhaps the most interesting is the one employing an autosegmental approach. According to Goldsmith (1979), /s/ consists of a supraglottal and a laryngeal tier. Seen in this light, aspiration occurs because the supraglottal tier is deleted and thus the laryngeal tier remains with its concomitant laryngeal features. That is why phonetically one hears aspiration. The word esto 'this' illustrates this point:

(1) e t o

$$\begin{array}{lcl} \left[\begin{array}{l}\text{Supralaryngeal}\\ \text{features}\end{array}\right] & \text{------------>} & \varnothing\ [\text{éhto}] \\ | & & \\ \left[\begin{array}{l}\text{Laryngeal}\\ \text{features}\end{array}\right] & & \end{array}$$

This analysis proves to be superior in explanatory power to all hypotheses proposed in the literature. In particular, it is more insightful than the most widely accepted, yet quite unrevealing, hypothesis of syllable-final /s/ aspiration or deletion.

The analysis in (1) offers a reasonable explanation of /s/ aspiration; nonetheless, other researchers have observed that there is a host of grammatical, phonological, and pragmatic variables that are statistically significant in determining whether /s/ would remain or whether aspiration or deletion would obtain (on this topic see Alba (1983), Cedergren (1973), Poplack (1979), Nuñez Cedeño (1979), among others, and the bibliographies found therein). I do not intend to renew discussions on these interesting theses. On the contrary, my purpose is to study the insertion of /s/ within the word, a phenomenon mostly done by semi-illiterates, illiterates and, to a lesser degree, by educated speakers, from Dominican Republic. In a pioneering study carried out in 1983, Terrell suggests that stylistic factors, and not phonological conditioning of any sort, trigger the occurrence of this segment. I will demonstrate that /s/ is not chaotically or arbitrarily inserted in the rhyme but that its insertion obeys the phonological constraints of the language and, furthermore, that syllable structure and stress placement restrict its occurrence. These two hypotheses are subject to what I call **property of epenthetic rules**, which aim to preserve structures generated by phonological rules.

Essentially, /s/ epenthesis in this dialect is a hypercorrective phenomenon. In general, speakers tend to hypercorrect, for instance, yo 'I', **disputa** 'dispute', or **loco** 'crazy' as the respective **yos**, **dipusta** or **losco**. In popular parlance, it is said that those speakers who articulate a final /s/, regardless of whether it is well-placed or not, tend to speak "fisno" < **fino** 'refined'. Historically, it has been shown that even the educated class tend to hypercorrect. Jiménez Sabater (1975:35) found in manuscripts written in 1790 and 1822 examples of "sastifación" < **satisfacción** 'satisfaction' showing, in effect, that aspiration or deletion of /s/ was an active process then, and that the scribe simply did not know where to insert the **s**.

Before probing in depth into /s/-epenthesis, let us first clarify for the present analysis the status of /s/ in the phonological inventory of our informants. In general, dialectologists agree that final /s/ in underlyingly present for literates and educated speakers. The controversy arises, however, when considering semi-illiterate or illiterate speakers. Terrell (1986:133) argues, for instance, that for illiterates, there is no phonetic evidence supporting the presence of /s/ in the rhyme. For him, phonetic forms in (2a) are identical to the lexical forms in (2b):

(2) a	b	
[en éto tiémpo]	/en eto tiempo/	"en estos tiempos" 'nowadays'
[el inglé]	/el ingle/	"el inglés" 'the Englishman'
[depué]	/depue/	"después" 'after'
[algúno lugáre]	/alguno lugare/	"algunos lugares" 'somewhere'

Others argue, however, that the singular-plural alternations found in **me - mes+e(s)** 'month - months', or **ve - vec+e(s)** 'time - times' show that lexically both are represented as /mes+e/ or /ves+e/. In other words, an /s/-deletion rule must be invoked in order to obtain the singular form **vez** 'time' which deletes the final /s/ (see Nuñez Cedeño 1979). A nonconcatenative analysis of plural formation proposed by Harris (1980) may lend support to Terrell's hypothesis. The alternations of what is known as double plural formation in Dominican Spanish, as in **mucháchose** < **muchachos** 'boys' or **gallínase** < **gallinas** 'hens' (the regular plural is normally formed by adding the plural allomorph **+s** to words ending in an unstressed vowel or **-es** to words ending in a consonant) suggest that speakers analyze plural forms with a VCV empty skeleton in which this sequence is interpreted as -Vse, VCe, or as an uninterpreted VCV, that is, as a [Ø]. The plural form **mujérese** 'women' is analyzed, grosso modo, /muher + VCV/. The final V is filled with **-e**, C with **-s**, and the penult V with **-e**. Monosyllabic forms undergo similar treatment, thus **mes** is represented for semi-literates as /me + VCV/ in which the penult V is associated with the class vowel **-e** and CV is filled with the allomorph **-se**, the output being **mese**.[1] Harris (1980) discusses this more extensively and motivates the VCV template and the default rules filling it.

The forms **e(scribir)** and **con(scribir)** are perhaps counterexamples against a final underlying /s/. The former shows an epenthetic **e-** disallowing the initial consonantal clusters #s(C)C; this is true at least for the aspirating dialects.

But these forms do not constitute hard-core evidence for the existence of /s/ in final position. It is quite possible that speakers simply have the respective /ekribir/ and /konkribir/. In fact, a speaker, aware of my knowledge of English asked me "yu pik ingli?" 'do you speak English?'. That the initial /s/ of "speak" was not pronounced shows that she lacks this segment in cluster initial position, for otherwise her question should have been "yu ehpik ingli?" in which case the epenthetic - e would have fulfilled its role of breaking up this type of cluster. What she did, then, was to adopt the linguistic borrowing and adapt it to her morphophonemic system, namely, to one in which there is no an underluying #s(C)C sequence. At this point it seems that the data favor the hypothesis of the inexistence of underlying /s/ in the rhyme for the speakers studied here. I will therefore assume this hypothesis until the opposite is demonstrated. I will further assume the organization of the syllable as discussed by Clements and Keyser (1983) for English, and Núñez Cedeño (1985, 1986), for Spanish. In the latter, almost any C(onsonant) of the phonological inventory, with a couple of exceptions, can be added to the syllabic prototype CV.

After this brief digression, consider now the data in (3):

(3) **Words out of context**

etús$pido	"estúpido"	'stupid'
is$no	"himno"	'hymn'
invis$tado	"invitado"	'guest'
dipus$ta	"disputa"	'dispute'
afes$tado	"afectado"	'presumptuous'
yus$ca	"yuca"	'yucca'
as$trá	"atrás"	'behind'
dos$tor	"doctor"	'doctor'
dedes$	"desde"	'since'
comos$	"como"	'like'
esquipo	"equipo"	'team'
revólvers$	"revólver"	'revolver'
gols$	"gol"	'goal'
cóndors$	"cóndors"	'condor'
des$viars$	"desviar"	'to turn away'
sobrans$	"sobran"	'lacking'

Words in context

ses # traen un aire	"se traen un aire 'they put on an air'
yos # tengo	"yo tengo" 'I have'
una nota altas#	"una nota alta" 'stoned'
me encargós # do pare	"me encargó dos pares" 'ha asked me to get two pairs'
se vas # a huir	"se van a huir" 'they are escaping'
luego aprendís # a tocar	"luego aprendí a tocar" 'then I learned to play the guitar'
él va a durars # cuatro año	"él va a durar cuatro año" 'he is spending four years'
aquís # vivimos yo y..	"aquí vivimos yo y.." 'here we live me and..'
yos # hago	"yo hago" 'I do'

From these data it is clear that /s/ insertion occurs in the rhyme, an observation which may lead us to believe that epenthesis is sensitive to a fully structured syllable and, furthermore, that it does not occur in any unstructured sequence of segments. The word **revólver** 'revolver', for instance, is lexically represented as a sequence of CVCVCCVC (here, each C and V represents a column of features in the usual sense). If epenthesis were to occur before syllable structure takes place, /s/ could randomly be assigned after the last C, yielding the correct **revólvers**, but it could also be inserted before or after the last V, in which case we would have the totally nonsensical and unpronounceable

***revólvser** or ***revólvesr**. It is unquestionable, then, that /s/ must find a fully formed syllable to which it may attach. The fact that /s/ is inserted at the end of the syllable is captured with rule (4):

(4) /s/ epenthesis

$$\emptyset \dashrightarrow s \;/ \underline{\qquad\quad} \sigma$$

Rule (4) accounts for all forms listed in (3). But its unquestioned acceptance goes beyond of what meets the eyes. In addition to accounting for the forms in (3), (4) has the ill-desired effect of generating anomalous forms. Consider, for instance, all the cases with vowels in hiatus, or the combinations of nondiphthongized vowels. In principle, nothing prevents the addition of /s/ to these forms and the production of nonexistent forms, as in (5):

(5)

te$a$tro 'theater'	Rule 4	----->	tes$a$tro	----->	*[tesátro]
grú$a 'crane'	"	----->	grus$a	----->	*[grúsa]
ca$os 'chaos'	"	----->	cas$os	----->	*[kásos]
yautí$a 'yautia'	"	----->	yautis$a	----->	*[yautísa]
po$e$ta 'poet'	"	----->	pos$e$ta	----->	*[poséta]
a$hí 'there'	"	----->	as$hí	----->	*[así]

The data themselves in (3) suggest a potential solution to the problem we are facing in (5). Note that, for instance, there is not a single form in which /s/ is followed by a vowel. The same holds true when insertion occurs across the word. A rule like (6) would resolve the issue and would account for the data in (3):

(6)

$$\emptyset \dashrightarrow s/\underline{\qquad\quad}\,\sigma \left\{ \begin{matrix} C \\ \# \end{matrix} \right\}$$

As a solution, (6) is unattractive on two counts: 1) it complicates the grammar and loses a valuable generalization, and 2) although observationally correct, it does not give us a satisfactory explanation as to why /s/ can be added before a consonant or at a word boundary but not before a vowel.

In (5), each input contains a well constructed syllable. Generally speaking, the syllabification rules specify exactly the type of consonants that will make up the onset. In this theory of syllable structure the coda is not inherently attached to the nucleus; it is obtained by rule. Given the sequence CVCCV (in (6)), a spontaneous syllabification would produce CVC$CV; the word **carta** 'letter' would be syllabified as **car$ta**. Exactly the same obtains for the hypercorrected **losco** < **loco** 'crazy' whose syllabification is **los$co**. The automatic syllabification for the sequence CVCVCV is CV$CV$CV, as illustrated with the word **ca$be$za** 'head'. Now, when /s/ is inserted into the sequence CVVCV creating the sequence CVCVCV, the insertion runs afoul of the principles of syllable construction by stipulating that the latter should be syllabified *CVCVCV. This is the case with the word **poeta** 'poet' which is syllabified **po$e$ta**. Upon adding /s/ to the first syllable, an erroneous syllabification is created, namely, ***pos$e$ta**, *CVCVCV. That is why the output *[poseta] is wrong, and the same is true for the remaining forms in (5). How do we go about capturing this fact and simultaneously maintaining the broad generalizations achieved by (4) without having to invoke the conditioning boundaries listed in (6)? This can be achieved by simply indicating that (4) obeys the structure and syllabification restrictions of the language. (4) is therefore constrained by the following property:

(7) Epenthesis is a structure-preserving rule.

The term "structure-preserving rule" is used here in the sense of Edmonds (1970), and it means that an epenthetic rule a) cannot create structures not generated by phonological rules[2], and b) cannot alter the phonological features of immediately adjacent segments.[3] With the first definition, there is no need for attaching conditions of applicability to (4), as exemplified by the contexts in (6); also, arbitrary and ad-hoc stipulations are eliminated. For instance, there is no need to say that in case of homonomy, (4) must be blocked. Notice that *[káso] < **caos** 'chaos' and +[así] < **ahí** 'over there' may be confused with the actual [káso] and [así].

The hypercorrection in the sentence **yos hago** 'I do' reveals that /s/-epenthesis only operates at the lexical level and that the word boundary # prevents /s/ from seeing the following segment.[4] That this must be the case is shown by the actual syllabification of the sentence [yós$á$go] which otherwise would be rejected by (7). A sample derivation is provided in (8):

(8)	/yo # ago/	
	yo$ # a$go	Syllabification
	yos$ # a$go	Rule 4
	[yós ágo]	Output

Before obtaining the final output, the above sentence may be resyllabified and ultimately we would have [yo-sá$go]; however, I have not been able to collect a single example of resyllabification. Even if the latter was to obtain, it is incontrovertible that (4) takes place at the lexical level[5] and that it is indifferent to what follows, for otherwise it would conflict with the predictions of (7). Therefore, (4) needs not be complicated with stipulation of syntactic boundaries.

That (4) meets the requirements of structural preservation is further supported by the data in (9):

(9)

a		b	
puedos, puesdo	"puedo" 'I can'	*cuersda	"cuerda" 'string'
cuabas, cuasba	"cuaba" 'Jamaica rosewood'	*cuarsto	"cuarto" 'room'
cielos, cieslo	"cielo" 'sky'	*ciensto	"ciento" 'hundred'

The syllables with diphthongs and checked by /s/ in (9a) are documented in most Spanish dialects. But notice what happens in (9b) which (4) could generate. When /s/ is inserted into these rhymes a principle of rhyme formation is violated, since it now exceeds its maximum length of three segments. The absence of (9b), then is attributable to the basic principles of rhyme formation and it is therefore in consonance with (7). Consider also the data in (10).

(10) a	b	
plástano, plátanos	*plátasno	"plátano" 'plantain'
cásñamo, cáñamos	*cáñasmo	"cáñamo" 'hemp rope'
párspado, párpados	*párpasdo	"párpado" 'eyelid'
ásguila, águilas	*águisla	"águila" 'eagle'
lástigo, látigos	*látisgo	"látigo" 'whip'

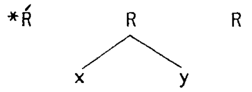
c		d	
fuscú, fucús	"fucú" 'bad luck'	císda, cádas	"cada" 'each'
caslamidá,	"calamidad" 'calamity'	múslo, múlos	"mulo" 'mule'
calasmidá, calamisdá, calamidás,		muschácho, mucháscho, mucháschos	"muchácho" 'boy'
rastán, ratáns	"ratán" 'rattan'	vérsde, vérdes	"verde" 'green'

(The acute marks have been added for illustrative purposes.)

All these are possible forms. I have not been able, however, to collect rhymes with two consonants. The existence of

flors, **sobrans**, and other listed in (3) and those of neighboring dialects, such as in **perspectiva** 'perspective' or **constante** 'constant' suggest that the absence of final consonantal clusters is merely due to an accidental gap of production. More important is stress placement in the word. (10a) and (10b) are antepenultimately stressed; (10c) are ultimately stressed, and (10d) are penultimately stressed. All are possible forms, except for (10b). Why are the forms in (10b) impossible? In terms of sequence of segments, they are all identical. As we might have suspected, stress is probably the culprit. It has been demonstrated by Harris (1983:109) and others that stress placement in Spanish is impossible if the penultimate syllable has a branching rhyme. The rule in (11) is being violated:

(11)



It is not surprising that speakers consistently avoid the forms in (10b) since the principle of syllable structure (11) will prevent bad forms from surfacing. Principle (7) is further corroborated.

Recall that (7) has a second definition which disallows a change in features of segments immediately adjacent to the inserted /s/. Notice what happens to the forms in (12a), possible but not documented, and to the wrong ones in (12a'):

(12) a	a'		
asmara o amaras	*amasra	"amara"	'would love'
caros	*casro	"caro"	'expensive'
toreros	*tosrero o *toresro	"torero"	'bullfighter'
aspura o apuras	*apusra	"apura"	'hurry up'
cuslo, culos		"culo"	'ass'
casna, canas		"cana"	'fan palm'
b	b'		
carresta o carretas	*casrreta	"carreta"	'cart'
asmarro o amarros	*amasrro	"amarro"	'I tie'
herreros	*hersrrero, *herresro	"herrero"	'blacksmith'

It seems that we are facing yet another problem, for we might have to specify the type of consonants that may follow /s/, and this, of course, means an additional stipulation to (4). But of all these consonants only [r̄] is to be avoided. We will have to reformulate (4) as in (13):

(13)

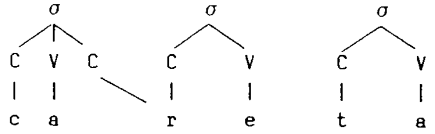
$$\emptyset \dashrightarrow s/\underline{\qquad}\,\sigma \quad \begin{array}{c} C \\ [-\text{vibration}] \end{array}$$

It is rather strange that (13) only excludes the simple flap /r/ from among all possible consonantal contexts, especially when the s+r sequence is widely attested in Spanish. Examples such as **Israel** 'Israel', **desriscar** 'to hurl', **desramar** 'to prune', and others support this observation. When /s/-epenthesis occurs there is a dramatic change in the underlying simple /r/. An /s/ added to the syllable **ma-** in the word **amara** 'would love', for instance, would trigger, by an independent rule, the change of underlying /r/ into a trill. /s/ insertion in this case causes a significant modification to the adjacent segment. Apparently, this is highly undesirable and it is to be avoided since there is no evidence of changes ever occurring in any other context. That adjacent segments do not change is independently motivated by **e-** epenthesis in Standard Spanish in word-initial -sC(C) clusters. After insertion nothing happens to the consonants adjactent to **e-**. Epenthesis is characterized by not altering immediately adjacent segments and accordingly the predictions of (7) hold.

Recall that in (3) there is little evidence for claiming that speakers generate syllables closed by two consonants within the word, though the hypercorrected **revólvers**, **flors**, and **canciónss** 'song' seem to indicate that it is possible to find hypercorrections like **parste** < **parte** 'part', **molsde** < **molde** 'mold' or **anste** < **ante** 'before'. This is not a spectacular revelation since these clusters are also found in general Spanish. The last two forms, however, seem to undermine principle (7) because, as it is known in Spanish, both /l/ and /n/ assimilate to the point of articulation of following consonants. With the addition of /s/, the already assimilated dentals [l̪] and [n̪] would have to once again assimilate to the /s/ alveolars. In accordance with this, both *[mól̪sde] and *[án̪ste] would be wrong. Now, the anomaly is created not because (7) is being contradicted on account of re-applying an assimilation rule to the same lexical form during the course of its derivation. This anomaly stems from the way in which epenthesis and assimilation interact with each other. The problem is simply avoided by ordering epenthesis before assimilation. The above forms support this ordering. This is further supported by the sentence **cobra[n]s plata** 'they collect money' which shows an alveolar [n]. Had assimilation occurred first, epenthesis would have met the output [coƀram plata] producing the degenerate *[coƀrams plata]. It is obvious that epenthesis is a late rule applying after syllabification but interacting with assimilation which is a postlexical rule. That is why **ca[n]spo** < **campo** 'country' is possible but not ***ca[m]spo**. Similarly, **co[n]sga** < **conga** 'conga' is right but not *co[ŋ]**sga**.[6]

The forms in (12b') suggest that (13), and not (4), is to be preferred since not only the former explains the anomaly of ***casra** < **cara** 'face' but of ***casrreta** as well. This would be true had we assumed an underlying distinction between /r/ and /r̄/. Harris (1983) has argued that such a distinction does not exist. Dominican Spanish further shows that this must be the case. In this dialect the standard trill is generally produced with a pre-aspiration followed by a simple tap. The word **perro**, for instance, is rendered as [péhro] which shows that the phonetic trill is indeed a sequence of taps as in /per$ro/. What is felt to be a pre-aspiration is a simple case of syllable-final /r/ aspiration (this phenomenon has been documented by Jiménez Sabater (1975:86)). I propose, then, that the forms in (12c) are analyzed with a heterosyllabic tap, as illustrated in (14):

(14)



'cart'

In (14) when syllable structure occurs, the initial syllable is syllabyfied as **car-**. Seen in this way, ***casrreta** is impossible because /s/ has not been inserted in a syllable-final position. Nor can it be inserted immediately after **car-** because epenthesis usually does not alter a geminate relationship of two to one (cf. McCarthy, forthcoming, Steriade 1982).[7]

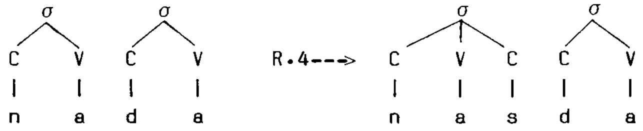
Throughout the discussion, I have argued for and motivated a rule of epenthesis in order to account for hypercorrections with s. But to this analysis an alternative could be offered, one in which /s/ is not inserted but it is already found underlyingly and meeting all wellformedness requirements that we have discussed so far. This interesting possibility is untenable when confronted with the recalcitrant facts of this dialect. Consider, for instance, those speakers that aspirate or delete syllable-final /s/. As indicated earlier, they tend to hypercorrect but their hypercorrections show a fully articulated /s/ suggesting that /s/ aspiration or deletion rules do not apply. Presumably the lexicon would have to be divided into two groups: a) one with hypercorrections whose forms are diacritically marked as not undergoing the above two rules, and b) the other in which the forms undergo aspiration or deletion. The point is that this would call for an enormous division of the lexicon since unhypercorrected forms alternate with hypercorrected ones. Both alternants somehow will have to be related at some point and this means that the grammar would be greatly complicated by unnecessarily overloading the speaker's lexicon. By Occam's razor, an insertion rule is preferable. This reasoning is also applicable to speakers lacking aspiration/deletion rules.

Let us stretch the above possibility further and suggest that speakers temporarily analyze a given form with an hypercorrected /s/, representing, grosso modo, what they believe to be a correct lexical item heard from a speaker in a formal situation. With this solution evidence will be needed for supporting the hypothesis that all processes or rules are also temporarily suppressed. In addition, it will have to be demonstrated that an underlying /s/ is totally unrelated to other rules. This is not the case, however. In Dominican, the following alternations are frequently found:

(15)	kuida$do	-	kuidao	"cuidado"	'careful'
	na$da	-	na	"nada"	'nothing'
	pa$ra	-	pa	"para"	'for'
	balbu$do	-	balbu(o)	"barbudo"	'bearded'
	komi$da	-	komía	"comida"	'food'
	a$buela	-	auela	"abuela"	'grandmother'
	a$gua	-	aua	"agua"	'water'
	to$do	-	to	"todo"	'all'

Now, let us suppose that speakers have heard the respective **pa** and **to**, why is it that they never hypercorrect by saying ***pas** or ***tos**? According to my hypothesis, this is due to the interaction of /s/ epenthesis and onset deletion. Take for instance (16):

(16)



In general, onset deletion occurs intervocalically to yield, for example, **na**. The insertion of /s/ bleeds onset deletion. But suppose that deletion precedes epenthesis, then we would have an intermediate derivation /na$a/. When rule (4) applies to this derivation the syllabification would produce the erroneous /nas$a/, disallowed by principle (7). This seems to be a reasonable explanation that cannot be achieved by the alternate hypothesis in which there is no principled way of knowing why speakers avoid the bad forms quoted above. These would have to be ad-hoc and arbitrarily listed as exceptions and, again, the lexicon will increase considerably, for it is known in this dialect that most past participles ending in **-ado**, and **-ido** may lose their intervocalic /d/. One final observation, the mere existence of final /s/ makes it all more impossible to falsify as a hypothesis. On purely theoretical grounds, a falsifiable and arguable solution (proposed here) is preferred to one which is not.

In summary, this analysis has shown that the hypercorrection of /s/ in Spanish is fully conditioned by structure preserving properties. These properties are the sole justification for the very existence of the simple rule of /s/ epenthesis.

NOTES

* I am thankful to Jim Harris, Jorge Guitart, and Tracy Terrell for their valuable suggestion and comments. The data gathered for this study were originally discussed in Terrell (1983, 1986). I have supplemented his data with spontaneous productions collected from native informants in a fieldwork research carried out in 1976.

[1] The same description applies to all monosyllabic adjectives and nouns closed by a final consonant. Yet, the noun **tos** 'cough' at first seems problematic because of its obvious relationship with the verb **toser** 'to cough', and this means that somehow the final /s/ will have to be accounted for, perhaps morphologically, as in $[[\text{tos}]\text{er}]_V$. In our case, this is not a thorny issue at all because, for semi-illiterates that noun underlyingly occurs without /s/, that is, as in /to/. I am suggesting that these alternants are independently listed in the lexicon but are semantically related by means of redundancy rules (cf. Riemsdijk and Williams 1986:38).

[2] It seems almost illogical, or even contradictonary, to match principle (7) with rule (4), for /s/-epenthesis is creating a non-existing structure, and we already know that there is no /s/ in syllable-final position. However, according to Nuñez Cedeño (1985), a single consonant can be attached to the syllabic nucleus, except for palatals, and the epenthetic /s/ meets the requirements of the rule, quoted here as (19):

> "(19) Final-C addition: add a final C to the
> syllabic prototype." (1985b:275)

/s/ insertion falls out from (19), therefore, nothing is violated.

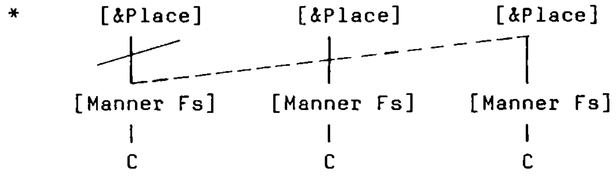
[3] It seems that property (7) is not exclusively a Spanish-specific rule. In fourteen different languages with epenthetic processes, all epenthetic rules obey structural wellformedness. The alterability of adjacent segments is not as general, however. This calls for more documentation since it seems to be the case that in one language the tendency is toward lenition (cf. McCarthy, forthcoming). In our case, the opposite holds, namely, a segment is strengthened, as we will see shortly.

[4] It so happens that for aspirating speakers with rule (1), and for those who have an additional rule of /s/ deletion, the application of (4) must follow these two rules. Lending support to this assertion is the fact that /s/ is never aspirated.

[5] /s/-epenthesis seems to occur only at the lexical level or perhaps at the very last stage of the word level. Consider, for instance, the word **novio** 'boy friend' which can be hypercorrected as **novios** or **nosvio**. Similarly, when the diminutive **-ecito** is added to the stem **noviecito** 'little

boy friend' is formed and after applying (4) **noviecisto**, or **noviecitos**, can obtain.

[6] Let us play the devil's advocate and argue that /s/-epenthesis follows nasal assimilation. Still, it would be impossible to derive co[ŋs]ga because epenthesis is crossing an intersecting line of an associated segment. Something like this would be created:



Therefore, epenthesis would violate the universal condition which stipulates that association lines cannot cross (cf. Steriade 1982:65).

[7] In this article McCarthy analyzes the predictions of OCP which prohibits at the melodic level the presence of adjacent identical consonants. (14) meets this requirement.

REFERENCES

Alba, O. 1983. "Función del acento en el proceso de elisión de la /s/ en la República Dominicana," in El español del Caribe, Orlando Alba, editor. República Dominicana: Prensa de la Universidad Católica Madre y Maestra.

Cedergren, H. 1973. The Interplay of Social and Linguistic Factors in Panama. Unpublished doctoral dissertation, Cornell University.

Clements, G.N. and S.J. Keyser. 1983. CV phonology: a generative theory of the syllable. Cambridge: MIT Press.

Edmonds, J. 1970. Root and Structure Preserving Transformations. MIT Doctoral dissertation.

Goldsmith, J. 1979. "Subsegmentals in Spanish phonology: an autosegmental approach," in Linguistic Symposium on Romance Languages: 9, William W. Cressey y Donna Jo Napoli, eds., Washington: Georgetown University Press.

Harris, J.W. 1980. "Nonconcatenative morphology and Spanish plurals." Journal of Linguistics Research 1:15-31.

Harris, J.W. 1983. Syllable structure and stress in Spanish: a nonlinear analysis. Cambridge: MIT Press.

Harris, J.W. 1986. "Epenthesis processes in Spanish," in Studies in Romance Linguistics, Carol Neidle and Rafael A. Nuñez Cedeño, eds., Holland: Foris Publications.

Jiménez Sabater, M.A. 1975. Más datos sobre el español de la República Dominicana. Santo Domingo: Publicaciones del InTec.

McCarthy, J.J. (forthcoming). "OCP effects: gemination and antigemination."

Núñez Cedeño, R.A. 1979. La fonología moderna y el español de Santo Domingo. Santo Domingo: Editorial Taller.

Núñez Cedeño, R.A. 1985. "On the three-tiered syllabic theory and its implications for Spanish," in Proceedings of the XIII Linguistic Symposium on Romance Languages, Larry King and Catherine Maley, eds., Holland: Benjamin Publications.

Núñez Cedeño, R.A. 1986. "Teoría de la organización silábica e implicaciones para el análisis del español caribeño," in Estudios sobre la fonología del español del Caribe, Rafael A. Núñez Cedeño, Iraset Páez Urdaneta and Jorge M. Guitart, eds., Caracas: La Casa de Bello, colección hispanoamericana de lingüística 2.

Poplack, S. 1979. Function and process in a variable phonology. Unpublished doctoral dissertation, University of Pennsylvania.

Riemsdijk, H. Van and E. Williams. 1986. Introduction to the Theory of Grammar, Cambridge: MIT Press.

Steriade, D. 1982. Greek Prosodies and the Nature of Syllabification. Unpublished MIT doctoral dissertation.

Terrell, T.D. 1983. "Relexificatión en el español dominicano: implicaciones para la educación," in El español del Caribe, Orlando Alba, Ed., República Dominicana: Prensa de la Universidad Católica Madre y Maestra.

Terrell, T.D. 1986. "La desaparición de /s/ posnuclear a nivel léxico en el habla dominicana," in Estudios sobre la fonología del español del Caribe, Rafael A. Núñez Cedeño, Iraset Páez Urdaneta and Jorge M. Guitart, eds., Caracas: La Casa de Bello, colección hispanoamericana de lingüística, 2.

Marianne Phinney

Resetting the Parameters: Acquiring Spanish as L2

1.0 Introduction

Research in language acquisition has often followed developments in linguistic theory quite closely. Beginning with Brown's group at Harvard (Brown 1973), acquisition researchers have drawn on linguistic theory for hypotheses about the acquisition process. In testing these hypotheses, new and rich insights have been reached about children's abilities to acquire language. More recently, the same approach is being used to examine the process of acquiring a second language.

Within current syntactic frameworks, Chomsky's Government-Binding framework (Chomsky 1981, 1982) is of particular interest to the acquisition theorist. The framework is designed to be relevant to acquisition, as various components of the grammer are specified to interact with the LAD. In particular, the notion of parameters, general properties of Core Grammar whose settings must be determined by input to the LAD, lead naturally to verification with data from acquisition.

In this paper, I will be concerned with one parameter which has been discussed by several theorists, and its role in the acquisition of Spanish as a second language. The Pro-Drop or Null Subject Parameter has been the subject of extensive discussion over the past few years (Chomsky 1981, 1982, Rizzi 1982, for example). The available data suggest that the ease of resetting these particular parameters in L2 acquisition can be partly predicted on the basis of the relative markedness of the parameters in the L1 and the L2.

2.0 The Null Subject Parameter in Spanish

It is well-established that Spanish, like many languages, does not require lexical subjects in many constructions. When the subject is personal and refers to someone previously mentioned in the discourse, the subject is normally omitted (1).[2]

(1) Trabajaré en la oficina de publicaciones de la universidad.
'I will work in the publications office of the university.'

In several constructions, it is impossible to have a lexical subject, as in the case of existentials (2a), statements about the weather (2b), and impersonal 3rd person referents (2c).

(2) a. Hay dos niños en ambas escenas.
'There are two boys in both scenes.'

b. Hace frío.
'It's cold.'

c. Es imposible.
'It's impossible.'

Other cases of empty subject positions in Spanish include the **se** "passive" (3a), verbs like **gustar**, and instances of postposed subjects (4).[3]

(3) a. Se venden hamacas.
'Hammocks are sold.'

b. Me gusta la playa mucho.
'I like the beach a lot.'

(4) Empiece la tentación.
'The temptation begins.'

It has often been suggested (Taraldsen 1978, Chomsky 1981, Picallo 1984, Suñer 1982) that these facts are associated with the presence of an extensive verbal inflectional system. Thus, given a reasonable discourse context, an empty subject is recoverable from the verbal morphology.

The traditional approach to the Null Subject parameter has treated all the empty subject positions in sentences like (1-3) similarly. Suñer (1982) proposed an interesting variation on this analysis for the parameter in Spanish. She argued that Spanish has three types of "empty" subjects which could be governed by the parameter: **pro**, which can be governed by [AG] ([+tense] in her analysis), [e], in the case of postposed subjects; and [0], in the case of existential and impersonal sentences like those in (5) and (6).

(5) Había mucha gente.
'There were many people.'

(6) Parece que está enfermo.
'(It) seems that (he) is sick.'

She notes that these sentences never have a pronominal subject, that the subject is semantically as well as phonologically empty, and argues that the subject cannot be pro or [e], using evidence from gerundive and infinitive constructions.

Suñer claims that the use of [0] subjects in these constructions is a variation along the [± obligatory subject] parameter she proposes. While it is generally assumed that empty subjects are one of the main manifestations of the parameter, in most cases it is assumed that sentences like (5) and (6) are parallel to the more common case of (7), where the empty subject is referential.

(7) Leemos muchos libros.
Read (1st pl.) many books.
'We read many books.'

Suñer's analysis thus presents a more complex view of the parameter; not only are there empty referential subjects, but there are also truly empty subjects which would then not come under the binding conditions. In the more standard analysis, (5-7) are accounted for by the same process. If Suñer's analysis is correct, the difference in the two types of constructions may be evidenced by different rates in L2 acquisition.[4]

3.0 The Place of Parameters in L2 Acquisition

Part of the framework of parameters entails an acquisition mechanism for setting them correctly for the particular language being acquired. Chomsky (1981) has assumed that all parameters are initially at the unmarked setting in Core Grammar. This concept can be roughly viewed as a linguistic expectation on the part of the child. Until the input data proves otherwise, the unmarked setting functions as a type of "working hypothesis" for the child.

Hyams (1983) presented considerable data to support the hypothesis that children initially assumed their L1 was a Pro-Drop language, and showed evidence of resetting the parameter in data from English children. She argued that children acquiring English took some time to set the parameter, while children acquiring Italian, a Pro-Drop language, appeared to produce subjectless sentences accurately (with correct verbal agreement) in the early stages of acquisition.

For adults acquiring a second language, other factors must be considered, specifically the relative setting of the particular parameter in the L1 and in the L2. An adult has already set the relevant parameter for her first language. In many cases, the parameter must be reset for the second language. Even in cases where the settings appear to be quite similar, new lexical and syntactic reflexes must be acquired, so L1-L2 transfer is not direct. The extensive literature on negative transfer (L1 interference) is testimony to the fact that L2 acquisition is not identical to L1 acquisition (even among children in immersion programs, some interference is evident). Some of this interference may be attributed to a need to reset a particular parameter.

It seems to be clear that L2 learners do not reset the parameter to its original unmarked state when L2 acquisition begins. White (1985) presents data from speakers of English learning French, neither a Pro-Drop language, which do not show evidence of the learners "returning" to the unmarked setting (null subjects) of the Pro-Drop parameter. Although the subjects make other errors, they do not produce subjectless sentences or accept them as grammatical in the target language.

The diagram in Table 1 shows the possibilities for acquisition of a particular parameter in terms of the amount of input required to trigger the resetting of the parameter (or, correspondingly, the relative 'ease' of acquisition). When the settings are similar (unmarked → unmarked or marked → marked), the amount of input needed would be that necessary to acquire the syntactic structure and lexical items. Because no resetting of the parameter would be required, acquisition of the affected constructions should be relatively rapid and the percentage of errors small.

		Setting of L1	
		Unmarked	Marked
Setting of L2	Unmarked	1	2
	Marked	3	1

Table 1. Ranking of ease of acquisition based on relative markedness of L1 and L2 parameter settings.

When the settings of the parameter differ in L1 and L2, a directional difference is predicted. Marked to unmarked (in this case English to Spanish) is predicted to require less input and thus be more rapid with fewer errors than unmarked to marked (Spanish to English).[5] In other words, resetting the Null Subject Parameter should be more difficult for Spanish-speaking students acquiring English than for English-speaking students acquiring Spanish.

A previous study (Phinney in press) with Puerto Rican students acquiring English suggests that a difference in parameter settings will affect the L2 acquisition process. An analysis of production data (compositions) shows that certain aspects of the parameter (the pleonastic subjects **it** and **there**, postposed subjects in relative clauses (Torrego 1984), and use of subject pronouns under discourse contexts which involve given information such as embedded clauses) take considerable time to acquire to criterion, and errors in these constructions still occur even in advanced students. Although there was an increase in the use of referential pronominal subjects as proficiency increased, they were still omitted by advanced students, particularly in embedded clauses. Non-referential **it** was omitted between 50% to 85% of the time, and there was an increase of omitted subjects in existential **there** constructions as proficiency increased. For some students, those aspects, particularly impersonal **it** subjects, are the only major errors that still persist (other than prepositions and articles).[6]

If it is assumed, following Chomsky (1981, 1982) and Rizzi (1982), that the Null Subject parameter is unmarked in Spanish and marked in English, then it appears from these data that resetting the parameter from unmarked to marked takes some time, and presumably requires considerable input. By the same token, English speakers acquiring Spanish will need to reset the parameter in the opposite direction, from marked to unmarked. While this is not equivalent to returning to the Core Grammar settings, with all that implies for first language acquisition, it is hypothesized that the period required for resetting will be shorter than in the other direction, and that the constructions which continue to cause ESL students problems will be relatively easy for SSL students.

4.0 The Study

4.1 Description

The study to be presented here examined the spontaneous written production of English-speaking college students learning Spanish. I am assuming that the demands of a timed composition task under classroom conditions will provide some evidence of their acquisition of Spanish rather than conscious rule learning (Dulay, Burt, and Krashen 1982). The texts were scored for use of various types of subject-tensed verb constructions.

4.2 Subjects

The subjects were students in the Spanish program at the University of Massachusetts at Amherst. Four levels were examined, first through fourth semester Spanish, with roughly twenty students in each level. The groups were intact class groups, placed by either an examination or by passing the previous course.

4.3 Method

The compositions consisted of assigned topics which varied between levels. They were written within a single class period without the use of dictionaries or the assistance of the instructor. Students were encouraged to be creative and show their mastery of the language.

4.4 Analysis

The texts were scored by clause (first or 'second'), agreement of the verb with the stated or intended subject, the type of subject (full noun phrase, pronoun, or empty) and whether the context prohibited a lexical subject (constructions like those in (2-3)). The data were grouped for each proficiency level. Within each category, percentages of usage were determined from the ratio of instances over obligatory contexts.

5.0 Results

Table 2 shows the breakdown of subject usage with and without subject-verb agreement at each level.[7] Note that empty referential subjects as in (1) predominate from the earliest stage, while use of obligatory empty subject contexts (2-3) increased (the difference between Level III and Level IV is significant, $p < .001$).

	Level			
Subject	I	II	III	IV
lexical NP +AG	14	11	209	122
pronoun +AG	35	32	55	50
empty subject +AG	224	89	134	94
lexical NP -AG	1	7	51	42
pronoun -AG	0	9	11	12
empty subject -AG	18	14	26	20
obligatory empty subject contexts	21	23	46	127
Total all contexts	313	167	532	467

Table 2. Breakdown of subject usage

Table 3 shows the percentage of use of referential subjects (excluding **hay** and **es** constructions). This information is also shown in the graph in Figure 1. Use of empty subjects decreases quite sharply in the higher levels, due to the increase in lexical NP subjects. This is partly due to the increased complexity of the topis discussed in the texts; as the discourse becomes more involved, it becomes necessary to refer to topics by name to avoid confusion. However, it should be noted that the use of empty subjects and full NP's does not yet approximate that of native speakers; in many cases, the referent had already been mentioned in the discourse and could have been referred to with **pro** rather than with a lexical phrase, as shown in (8).

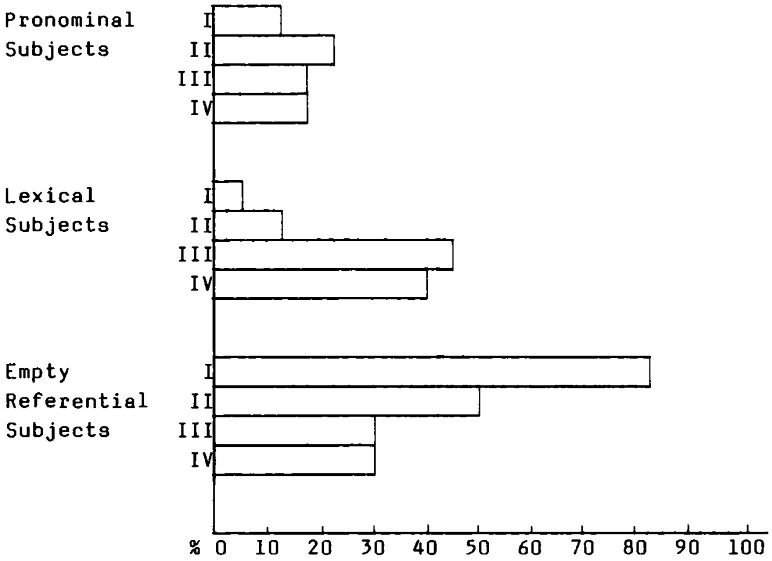
(8) Es verdad tambien en "La Vuelta...", pero en "La Vuelta" uno de los niños ha caido en la remanso y **ese niño** necesitó de ayudo de Roque Carpio.

'It is true also in "La Vuelta...", but in "La Vuelta", one of the boys has fallen in the backwater and **that boy** needed the help of Roque Carpio.'

	Level			
Subject	I	II	III	IV
lexical NP +AG	4	8	43	36
pronoun +AG	12	22	11	15
empty subject +AG	77	48	28	28
lexical NP +AG	1	5	10	12
pronoun -AG	0	6	2	5
empty subject -AG	6	10	5	6

Table 3. Percentage of referential subject usage

Figure 1. % of subject usage with correct verbal agreement English speakers acquiring Spanish



The use of contexts in which a null subject is obligatory, as in (2), is low compared to the total number.[8] Nevertheless, a closer examination of these sentences proves interesting. In table 4, the various null subject constructions are broken down by group. Although the usage of these constructions is very much dependent on the topic of the discourse, it can be seen that constructions with **gustar** are very frequent except in Group III, where the topic did not lead to using **gustar**-type verbs; however, there is a corresponding increase in sentences with **hay**, and some students seemed quite fond of it.[9] The use of **hace**, as in (2b), is extremely rare, as are postposed subjects and impersonal **se** constructions. This may be indicative of the topics of discourse; on the other hand, their rarity may be a manifestation of avoidance, and thus indicate delayed acquisition.

	Level			
Verb type	I	II	III	IV
hay	14	9	50	17
es	14	17	4	9
es (referential)	10	22	39	13
gustar	43	52	4	57
hace	17	0	2	3

Table 4. Percentage breakdown of obligatory empty subject contexts

In considering the analysis of the Null Subject parameter proposed by Suñer (1982), the data in Tables 5 and 6 are crucial. Note that the attested frequency of empty subjects with these obligatory null constructions is extremely high, as shown in Table 5, as is the frequency of correct usage (Table 6). The decrease in Levels III and IV can be attributed to the errors of a few individuals in each group, indicating that the resetting of the parameter may be subject to individual variation. However, when compared to the ESL students from the previous study (Phinney in press), it is clear that the students learning Spanish have reset at least part of the parameter quite early in the acquisition process.

	Level			
Verb type	I	II	III	IV
hay	100	100	100	86
es	100	100	100	75
es (referential)	100*	80	78	77
gustar	100	83	50*	85
hace	100	-	100*	75

* small N

Table 5. Percentage of Ø subject usage (± AG, ± correct verb)

Verb type	Level I	 II	 III	 IV
hay	67	100	96	67
es	67	100	100	75
es (referential)	100*	80	78	71
gustar	100	75	0*	73
hace	100	-	100*	0

* small N

Table 6. Percentage of correct usage of obligatory empty subject contexts

Despite this early success, there is still evidence of persistent interference for these students. The high use of lexical NP's in contexts where the discourse would normally require at most a pronoun and more likely an empty subject indicates that the parameter is not completely reset.[10] Indeed, awareness of the rhetorical factors involved in subject usage may be part of the process.

Figures (2-4) show the use of subjects by clause, first or second. While preliminary, this analysis indicates that awareness of discourse considerations begins to appear at the upper levels. NP usage is considerably higher in the first clause than in the second. Pronoun usage is slightly higher in the second clause. Empty subjects are much more frequent in the second clause, indicating that the students are beginning to reflect normal discourse patterns. Although by native speaker standards students still overuse subjects, they appear to be working out the relationship between the parameter and the context of the discourse.

Figure 2. Percentage of NP subjects by clause

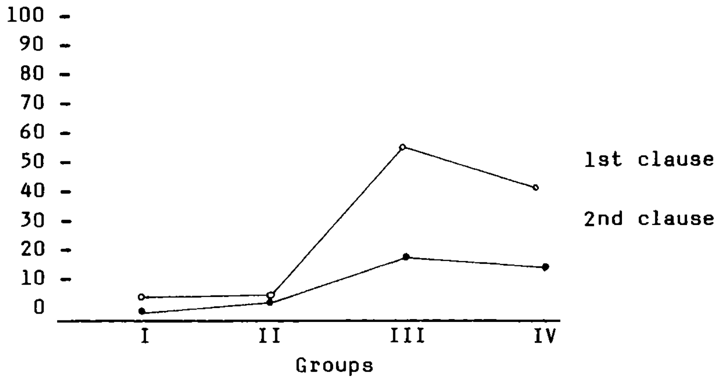


Figure 3. Percentage of pronoun subjects by clause

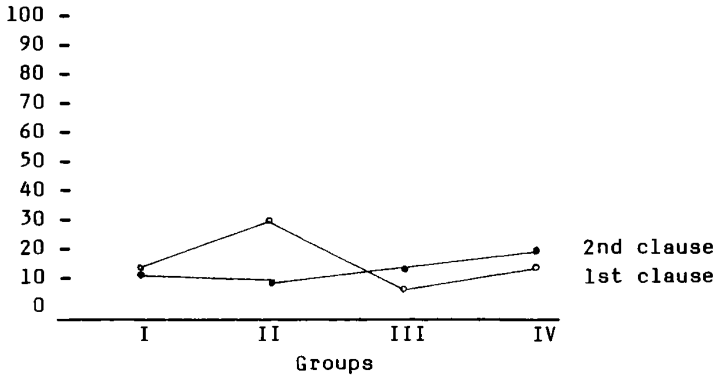
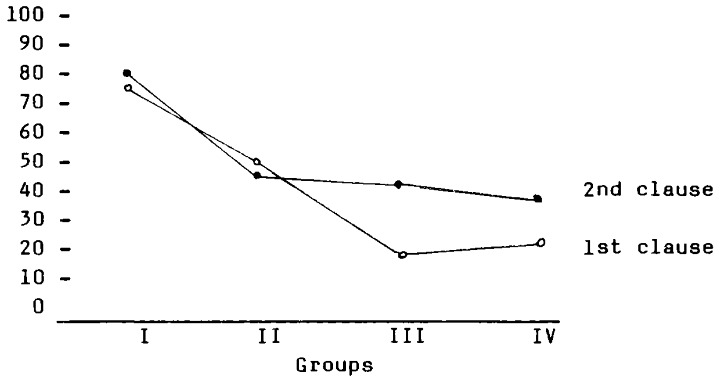


Figure 4. Percentage of empty subjects by clause



6.0 Conclusion

To summarize briefly, the English speaking students learning Spanish studies here show evidence of resetting the Null Subject parameter quite early in the acquisition process, compared to their Spanish speaking counterparts learning English. In addition, the data seem to indicate that not all aspects of the parameter are incorporated into the grammar at the same time. The data also indicate that there is a difference in the acquisition of the obligatory empty subject constructions such as **hay** and **es** as opposed to empty referential subjects. Sentences with obligatory empty subjects, although they occur with less frequency, are more often correct, and are almost never used with overt subjects. This is compatible with Suñer's (1982) treatment of these items as a distinct aspect of the parameter. In addition, the paucity of certain constructions in spontaneous production may indicate avoidance of difficult constructions.

Finally, it appears that the apparent decrease in empty subjects in higher proficiency levels, which appears to be "backsliding", can be attributed to several factors. The complexity of topics and the students' corresponding increased linguistic sophistication combine to increase the overall complexity of the task, leading to increased use of lexical NP subjects to make the reference in the discourse clearer. More importantly, the data show a growing awareness of discourse conditions on the use of empty subjects. This awareness appears to affect not the resetting of the parameter but the real-world production of the constructions involved. Overall, the data support the predictions of the parameterized model as a theoretical approach to second as well as first language acquisition.

NOTES

[1] I would like to thank Dr. Harlan Sturm and Dr. Marie Sheppard of the Department of Spanish and Portuguese, University of Massachusetts, for providing me with the data for this study.

[2] Example sentences (1-4) are taken from the compositions written by the Spanish students.

[3] There has been some debate about the status of these constructions with respect to the Null Subject Parameter. In each of these constructions, the verb agrees in number with the following noun phrase, and the noun following the verb appears to have been postposed from the preverbal subject position. Chomsky (1982) assumes the base form to be that shown in (3a). Due to the low occurence of these constructions in the data base analyzed here, the details of the syntactic analysis will not affect the interpretation of the data, so I will continue to assume that the absence of lexical material in the subject position is related to the parameter, although the empty category in the subject position may not be identical to the empty category in (1).

[4] There should also, of course, be a difference in L1 acquisition; however, there is little available data on the early stages of the acquisition of these constructions in Spanish.

[5] In terms of this study, speculation about levels of input remains just that. Although years of study might be roughly equivalent to overall L2 input, there is no way to obtain evidence on the amount of exposure to particular constructions these learners have had.

[6] A pilot study of students' judgments of sentences with missing subjects conducted with the same population showed a similar pattern, while the effect of the clausal context was even stronger (Phinney 1984).

[7] Agreement was defined as using the appropriate personal ending for the intended or expressed subject. Errors in tense usage only were not counted as errors of agreement.

[8] The increase in Level IV stems from the topic, "Gente y cosas que no me gustan", which naturally produced a high percentage of **gustar** constructions.

[9] For the purposes of this study, the impersonal **es** constructions have been separated from cases in which the "it" subject could be taken as referential, referring back to a specific noun in the discourse context.

[10] At this stage of analysis, the data have not been scored for correct anaphoric use of pronouns (Luján, personal communication). As the ensuing discussion will make clear, the texts indicate that the anaphoric conditions on pronoun use in Spanish is one aspect of the parameter that has not been acquired.

REFERENCES

Brown, R. 1973. A first language. Cambridge, MA: Harvard University Press.

Chomsky, N. 1981. Lectures on government and binding. Dordrecht: Foris.

Chomsky, N. 1982. Some concepts and consequences of the theory of government and binding. Cambridge, MA: MIT Press.

Dulay, H., M. Burt, and S. Krashen. 1982. Language two. New York: Oxford University Press.

Hyams, N.S. 1983. The acquisition of parameterized grammars. Unpublished Ph.D. dissertation, CUNY.

Phinney, M. 1984. Parameter setting and interference in L2 acquisition. Presented at the annual meeting of the Linguistic Society of America, Baltimore, Maryland.

Phinney, M. In press. The Pro-Drop parameter in second language acquisition. Parameter Setting, ed. by Thomas Roeper and Edwin Williams. Dordrecht: D. Reidel.

Picallo, M.C. 1984. The Infl node and the Null Subject parameter. Linguistic Inquiry 15(1).75-102.

Rizzi, L. 1982. WH Movement, negation, and the Pro-Drop parameter. Issues in Italian Syntax. Dordrecht: Foris.

Suñer, M. 1982. On null subjects. Linguistic Analysis 9(1).55-78.

Taraldsen, K.T. 1978. On the NIC, vacuous application and the That-trace filter. Mimeographed manuscript, MIT. Distributed by the Indiana University Linguistics Club, Bloomington, Indiana.

Torrego, E. 1984. On inversion in Spanish and some of its effects. Linguistic Inquiry 15(1).103-29.

White, L. 1985. The "Pro-Drop" parameter in adult second language acquisition. Language Learning 35(1).47-62.

Yves Roberge

Clitic-Chains and the Definiteness Requirement in Doubling Constructions

1. Introduction

It is often assumed that subject clitics (henceforth, SCl) are base-generated in AGR of INFL; cf. Bouchard (1982), Brandi and Cordin (1981), Everett (1985), Jaeggli (1984) and Sportiche (1983). In Roberge (1986a,b), this approach is motivated on the basis of the crosslinguistic observation that tensed sentences in SCl-languages require the presence of either a lexical NP subject or a SCl (or both). Contrary to languages where subject pronouns are generated in [NP,IP] position, SCl-languages are never null subject languages where this property is understood as allowing the complete absence of a lexical subject, be it a full lexical NP or a pronoun.

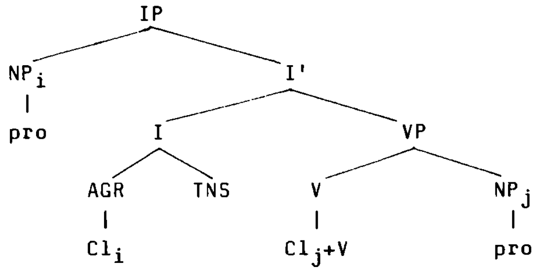
It is further argued that this generalization can be derived from a theory of pro similar to the one developed in Rizzi (1985) if one assumes that the SCl in INFL licenses pro in [NP,IP]. The obligatory presence of SCl when there is no lexical subject is thus accounted for since it is the licensing element for the empty category pro - like agreement markers in, say, Italian.

This discussion can be extended to cover the cases of object clitics (henceforth, OCl) which are analyzed as licensing elements for pro in the internal argument position they are related to.

Under these assumptions, a French sentence (1a) with a SCl and an OCl has the structure in (1b).[1]

(1) a. Il l' a vu.
SCl OCl has seen.
'He has seen her/it/him.'

b.



1.1. One of the predictions made by the preceding discussion concerns the clitic-doubling phenomenon. If, as we have seen, the relationship between a clitic and the argument it is related to is the same for internal and external arguments, we would expect clitic-doubling to be possible for lexical subjects. In other words, there seems to be no obvious reason why **subject doubling** (henceforth, SD) should not exist in parallel with **object doubling** (henceforth, OD). This prediction is borne out: certain languages with SCls exhibit SD. Some examples are provided in (2)-(6).

(2) Colloquial French

a. Le cuisinier il a fait cuire le poulet.
the chef he has made to-cook the chicken
'The chef cooked the chicken.'

b. Marie elle l' a mangé.
she it has eaten
'Marie ate it.'

(3) Trentino (Brandi and Cordin 1981)

a. Le putele le ven.
the girls they come
'The girls come.'

b. La taola l' ei costada masa.
the table it has cost a-lot
'The table cost a lot.'

(4) Fiorentino (Brandi and Cordin 1981)

a. Il ragazzo gl'é venuto.
the boy he has come
'The boy came.'

b. La tavola la é costata un monte.
the table it has cost a lot
'The table cost a lot.'

(5) Modenese (Safir 1985)

I to fio i magnen.
your children they eat
'Your children eat.'

(6) Pirahã (Everett 1985)

Kóxoí hi ti xibáobá.
he me hit
'Kóxoí hit me.'

Detailed analyses of the construction in these languages can be found in the references above and an account of the Colloquial Fr. data is proposed in Roberge (1986a) where arguments can be found against an analysis of the phenomenon that would treat the lexical NP subjects in (2) as left-dislocated NPs.[2]

1.2. In this paper we will be concerned with a property often associated with OD and SD in certain languages, namely the impossibility of extracting the NP argument in these constructions. In the following section, five Romance languages or dialects are examined and it is argued that the (im)possibility of extraction out of doubling constructions must be a parametrized property. In the third section, an account in terms of Clitic-chain is proposed.

2. Extraction

2.1. River Plate Spanish

Jeaggli (1982) points out that a clitic doubled object NP cannot be extracted by Wh-movement in River Plate Sp. as shown in (7).

(7) a. Lo$_i$ vimos a Guille$_i$.
him saw+1p to
'We saw Guille.'

b. *A quién$_i$ lo$_i$ vimos t$_i$?
to whom him saw+1p

c. A quién$_i$ vimos t$_i$?

Furthermore, extractions taking place in Logical Form (LF) are also barred. The quantified NP objects in the ungrammatical sentences in (8) undergo Quantifier Raising (QR) in LF; cf. May (1977), Hornstein (1984).

(8) a. Las$_i$ ví a todas las chicas$_i$.
them saw+1s to all the girls
'I saw all the girls.'

b. *Las$_i$ encontré a algunas mujeres$_i$.
them met+1s to some women
'I met some women.'

c. *Lo$_i$ ví a un chico$_i$.
him a boy
'I saw a boy.'

2.2. Colloquial French

Subject doubling in Colloquial Fr. obeys the same type of restriction, i.e. extraction from subject position is impossible when there is a coreferent SCl in the same clause. This is illustrated in (9).

(9) a. *Qui$_i$ il$_i$ est allé?
who he is gone
'Who went?'

b. *[$_{CP}$Wh$_i$ [$_{IP}$ t$_i$ Cl$_i$+V ...]]

c. Qui est allé?

There is also, parallel to the River Plate Sp. example in (8c), a definiteness requirement on the lexical NP associated with a SCl.

(10) a. La fille elle va chanter.
the girl she go to-sing
'The girl will sing.'

b. *Une fille elle va chanter.
a
'A girl will sing.'

c. Une fille va chanter.

The string in (11b) must therefore be excluded.

(11) a. $[_{IP}\ \underset{+def}{NP_i}\ \ Cl_i{+}V\ ...]$ = (10a)

b. $[_{IP}\ \underset{-def}{NP_i}\ \ Cl_i{+}V\ ...]$ = (10b)

If it is true that indefinite NPs are quantified expressions and, as such, must undergo QR at LF then, in (11b), the NP is extracted for scope interpretation requirements. The string is therefore:

(12) $*NP_i\ [_{IP}\ t_i\ \ Cl_i{+}V\ ...]$

which is very similar to the one in (9b). What we want to suggest here is that (10b) is excluded for the same reason (9a) is. As pointed out in Jaeggli (1982), this reason can be viewed as the impossibility of extracting a clitic doubled NP. Furthermore, the fact that QR as well as Wh-movement is excluded suggests that the constraint responsible for blocking movement holds at S-structure as well as LF, in Colloquial Fr. and River Plate Sp. at least.

2.3. Trentino[3]

The Trentino facts have been presented in Brandi and Cordin (1981) and different analysis have since been proposed; cf. Bouchard (1982), Jaeggli (1984), and Safir (1985).

In Trentino, SCls are obligatory (except for 1s, 1p, and 2p for which there are no pronominal clitics) either in isolation or when a lexical preverbal NP is present, as in (13), but are excluded in the case of free inversion, as in (14).

(13) a. *(El) magna.
he eats

b. El Mario *(el) magna.

c. *(Le) ven.
they come

d. Le putele *(le) ven.
the girls

(14) a. (*El) magna el Mario.

b. (*Le) ven le putele.

c. E' vegnú 'na putela.
has come a girl
'A girl came.'

The sentences in (15) show that there is no definiteness requirement in SD and that quantifiers can freely cooccur with coreferent SCls.

(15) a. Qualche putel l' é vegnú.
some boys they have come

b. Una putela l' ei vegnuda.
a girl she has come

It is thus plausible to assume that extraction of a preverbal clitic doubled NP subject is possible in this Italian dialect. Now, the facts are not so simple and a qualification is in order here. In effect, Wh-movement cannot take place in this construction as shown in (16) to (19) where a SCl cannot surface after subject extraction.

(16) a. Quante putele è na via?
how+many girls are gone away
'How many girls have left?'

b. *Quante putele èle nade via?

c. *Quante putele le e nade via?

(17) a. Chi ha magna?
'Who has eaten?'

b. *Chi **halo** magna?

c. *Chi **el** ha magna?

(18) a. Quante putele penset che sìa vegnú?
think+2s that are come
'How many girls do you think have come?'

b. *Quante putele penset che **le** sìa vegnude?

(19) a. Chi penset che magna?
who think+2s that eats
'Who do you think is eating?'

b. *Chi penset che **el** magna?

It must thus be assumed that whatever constraint is violated in (16) to (19) crucially holds at S-structure but not at LF in order to allow for (15).

2.4. Fiorentino

In this other Italian dialect the facts are basically the same as in Trentino, i.e. SCls are obligatory either in isolation or when a lexical preverbal NP subject is present as shown in (20).

(20) Qualche ragazzo gl'é venuto.
some boy he has come
'Some boy came.'

One of the major differences between Fiorentino and Trentino concerns the presence of a preverbal impersonal clitic in free inversion constructions. Contrast (21) with (14c).

(21) a. Gl' é venuto una ragazza.
Cl(imp) has come a girl

b. Gl'ha telefonato una ragazza.
has telephoned

c. Gl'ha portato la torta una ragazza.
has brought the cake

We have no data involving Wh-movement from subject doubling in Fiorentino. The analysis proposed here will nevertheless make some predictions which would have to be tested. Clearly though, extraction can take place at LF as is the case in Trentino.

2.5. Standard French

The situation in Standard Fr. is the following: there is no free inversion (22), subject doubling is barred (23), and there can be no Wh-movement out of the subject position when a coreferent SCl is present (24). Note that the latter property can be interpreted as a consequence of the non-availability of SD.

(22) a. *Est venu Pierre.
is come
'Pierre has come.'

b. Pierre est venu.

(23) a. *Pierre il est venu.
he

b. *Les trois filles elles ont bu.
the three girls they have drunk

(24) a. *Qui il est venu?
'Who has come?'

b. Qui est venu?

c. *Combien de filles dis-tu qu' elles ont
how+many of girls say you that they have

bu du thé?
drunk of tea

'How many girls did you say drank tea?'

2.6. Summary

Let us set aside the question of knowing why a given language exhibits SD but not OD or vice versa. We can now establish some basic ways in which the languages under study here differ.

(25) a. +/- free inversion[4]
b. +/- subject doubling
c. +/- object doubling
d. +/- Wh mvt of a clitic doubled NP
e. +/- LF mvt of a clitic doubled NP

Any account of the relationship between clitics and argument positions will have to leave open:

A) the option of having doubling at all;
B) the option of extracting out of doubling in syntax;
C) the option of extracting out of doubling in LF.

3. Previous accounts

To my knowledge, none of the analyses already available can handle the three options above in a satisfactory manner. We will review three of them.[5]

3.1. Jaeggli 1982

In this analysis proposed to account for the impossibility of extracting a clitic doubled NP in object position in River Plate Sp., it is assumed that an OCl affixed to V absorbs both government and Case assigning properties of V. The empty category in the ungoverned object position is PRO.

This analysis makes strong predictions with respect to extraction. In effect, the Empty Category Principle (ECP, Chomsky 1981) will rule out any occurrence of a trace in a clitic doubled object position since this trace would not be properly governed. Note that with the ECP applying at LF, extractions in syntax as well as in LF are blocked in the same way.

3.2. Borer 1984

For Borer, the definition of proper government includes a condition on Case agreement between the governor and the governee. She assumes further that a clitic properly governs the argument position it is related to if, among other things, the clitic and this position bear the same Case. Since a clitic is a spell-out of Case features on a head it follows that the argument position receives Case from a so-called dummy Case-marker (Kayne's Generalization). Consider now the sentences below from River Plate Sp. (26a) and Romanian (26b).

(26) a. *A quién$_i$ lo$_i$ vimos **t**$_i$?
DAT whom him(ACC) saw+1p DAT

b. Pe care$_i$ credeai că am văzut-o$_i$
ACC which+one thought+2p that have seen it(ACC)

t$_i$?
ACC

In (26a), the preposition **a** assigns dative Case to the object **quién**. After Wh-extraction, the trace is also marked dative. The OCl **lo** is accusative. There is therefore Case mismatch between the governor and the governee and the ECP rules out the sentence since the trace is not properly governed.

In (26b) the situation is quite different according to Borer. The preposition **pe** assigns accusative Case, the trace is thus marked ACC but the OCl **o** is also ACC. The Case matching requirement being met, the clitic can properly govern the trace making extraction possible.

This type of analysis relies heavily on the use of dummy Case-markers to explain language variation. This causes a problem in SD constructions where this particular type of rescuing device is not appealed to; cf. (2)-(6).[6] In these examples there is no reason to assume that the clitic doubled NP and the clitic do not bear the same (nominative) Case. How then are we to explain the impossible extractions either in syntax or at LF?

3.3. Jaeggli 1984

This analysis constitutes one of the recent attempts at explaining the facts about extraction and inversion. Jaeggli presents an account of the impossibility of certain syntactic extractions in which no appeal is made to the ECP. His approach is motivated by the fact that Wh-movement from preverbal position to the nearest COMP is blocked in Italian (27b). Extraction must take place form the postverbal position as in (27a).

(27) a. Quante ne sono cadute?
how+many of+them are fallen
'How many of them have fallen?'

b. *Quante sono cadute?

If it is assumed that postverbal extraction is forced by the ECP, then there is no reason why it should apply in (27). The S-structure for (27b) is found in (28).

(28) $[_{S'} Wh_i \ [_S \ t_i$ INFL VP]]

The trace is properly governed by the Wh-phrase in COMP, yet the structure is ungrammatical.

Jaeggli adopts a contextual definition of empty categories according to which the empty category in (28) is actually pro in Italian since rich agreement governs it. We thus have (29) instead of (28).

(29) $[_{S'} Wh_i \ [_S \ pro_i$ INFL VP]]

Jaeggli further assumes the following:

(30) pro cannot function as a variable if locally bound by an operator.

Hence, if extraction in (27) takes place from preverbal position then pro is interpreted as a variable. The sentence is excluded as in (27b) according to (30). If extraction is from postverbal position then the pro subject is not coindexed with the operator and the sentence is grammatical as in (27a). Jaeggli claims that this approach can be extended to certain facts in Trentino since, as we have seen, preverbal extraction is impossible in most cases of Wh-movement.

The major problem with this claim concerns the definiteness requirement in doubling constructions. As shown in section 2.3, there seems to be no such requirement in Trentino. Since an indefinite NP undergoes QR at LF, then there is in fact extraction from preverbal position which, in Jaeggli's view of empty categories, leaves a pro locally bound by an operator.[7]

3.4. Summary

We have seen that two analyses in terms of the ECP (Jaeggli 1982, Borer 1984) fail to make a necessary distinction between syntactic extraction and LF extraction out of doubling constructions. Jaeggli's (1984) account, which does not rely on the ECP, still cannot handle the fact that a particular grammar can allow LF extractions while blocking syntactic extractions in the same construction.

4. Proposal

Let us return to the discussion in section 1 where it is assumed that a clitic must be related to an argument position. Borer (1984:118) states essentially the same condition when she mentions that 'the configuration $[Cl_i\ [e_i]]$ is a discontinuous pronominal element'. Let us push this intuition further in assuming the following well-formedness condition on the occurrence of clitics.

(31) A clitic must form a Clitic-chain (C-chain).[8]

This condition can be derived from the interaction of various principles and theories of Universal Grammar. The Extended Projection Principle (Chomsky 1982) requires the presence of an external argument position and other argument positions in a structure where a verb θ-marks them. These positions can be empty or lexically realized. Various conditions regulate the presence of empty categories. If the analysis presented in Roberge (1986a,b) is correct then the theory of pro accepts a coindexed clitic as its licensing element deriving therefore the condition in (31).[9]
A C-chain is defined as in (32).

(32) C-chain (definition)
α and β form a C-chain if:
i. α and β are coindexed;
ii. α governs β;
iii. α is a spell-out of case features;
iv. β is locally free.

Condition (iv) on the formation of a C-chain will block any extraction out of the argument position (β) related to a clitic (α) since β would then be locally A'-bound. The same condition rules out (lexical or empty) anaphors in doubling constructions since they are locally A-bound. Pronominals, which are locally free, and R-expressions, which are free, are not ruled out.[10]

In the next section, we will see how condition (31) and the definition in (32) can account for the properties presented above in different languages. The level at which this condition holds will play a crucial role whereas a supplementary (parametrized) condition is shown to be active in languages like Standard Fr. where clitic doubling is not attested.

5. Applications

In River Plate Sp. there is no extraction (QR or Wh-movement) out of object doubling constructions. Condition (31) therefore holds at LF and blocks the appearance of a variable associated with a clitic. The trace left by Wh-movement or QR violates (iv) in (32) and the sentence is ruled out.

The properties in Colloquial Fr. with respect to doubling are the same as in River Plate Sp. except that doubling in Colloquial Fr. is limited to the subject. The analysis is thus the same, i.e. condition (31) holds at LF.

In Trentino and Fiorentino the situation is different; LF extractions out of subject doubling are allowed whereas Wh-extractions are barred. This can be taken care of if we assume that condition (31) holds only at S-structure where it blocks the variable left by Wh-movement whereas a variable resulting from the application of QR will not be filtered out.

Now with respect to Jaeggli's (1984) conclusion that Wh-movement of a subject in Trentino takes place form postverbal position, we are left with two plausible hypotheses in interpreting the data.

A. Extraction from postverbal position is forced because condition (31) holds at S-structure and would otherwise be violated.

B. Extraction takes place from preverbal position but because of (31) a clitic cannot be present at S-structure together with the variable.

The test case that would provide us with evidence for A or B is an interrogative in which the subject is questioned in Fiorentino. Recall that in the inverted constructions in Fiorentino the subject appears postverbally but an impersonal clitic appears preverbally. Hypothesis A predicts that such an interrogative, corresponding to (33a), would have the form in (34a) not the one in (34b). Hypothesis B, on the other hand, predicts that (34b) is the grammatical output.

(33) a. Una ragazza l'é venuta.
b. Gl'é venuto una ragazza.
c. *L'é venuta una ragazza.

(34) a. Chi gl'é venuto?
$[_{S'} Wh_i\ [_S\ pro_j\ \ Cl_j{+}V\ \ t_i]]$

b. Chi é venuta?
$[_{S'} Wh_i\ \ [_S\ \ t_i\ \ \ V\ ]]$

Dieter Wanner pointed out to me that (34a) is the grammatical interrogative; hypothesis A is therefore the right one, which reinforces Jaeggli's conclusion. What is important here is that both hypotheses are compatible with our suggestions.

We now turn to Standard Fr. where clitic doubling does not exist for either subjects or objects. The definition of C-chain is designed to allow a lexical R-expression to be doubled by a clitic. Clearly this possibility must be blocked in Standard Fr. We propose to add a further condition to (32) which is a parametrized option to be fixed positively or negatively.

(35) v. β is [+pronominal]

This condition is present in the grammar of the languages where clitic doubling is impossible and blocks clitic doubled R-expressions which are [-anaphor,+pronominal].[11] Needless to say, (31) holds at LF in Standard Fr.

6. Conclusion

In this paper, we have argued that no previous account of the (im)possibility of extraction of OD constructions could handle the facts about extraction out of SD in Colloquial Fr., Trentino en Fiorentino (definiteness requirement and Wh-movement). We have pointed out that a unified analysis of these facts has to adequately permit or block:

A) the option of having doubling at all;
B) the option of extracting out of doubling in syntax;
C) the option of extracting out of doubling in LF.

The condition in (31) on the appearance of clitics, namely that they form a C-chain as defined in (32), can account for extractions in LF and/or in syntax depending on where this condition holds (in LF or at S-structure) in a particular grammar.

As far as the option of having clitic doubling is concerned, a parametrized condition on C-chains making reference to the pronominal or non-pronominal status of the argument with which the clitic is coindexed, makes it possible to block the presence of a lexical R-expression in such an argument position. In other words, the definition of C-chain given in (32) allows for doubling whereas the addition of the condition in (35) blocks this possibility.

NOTES

* I am indebted to Michael Rochemont for helpful suggestions and comments. Thanks also to Osvaldo Jaeggli, Diane Massam and Marie-Thérèse Vinet. This work has been funded in part by a grant from the SSHRC (Canada).

[1] Kayne (1984) argues that French subject pronouns are base-generated in [NP,IP] position and are then cliticized in the phonological component of the grammar in non-inversion constructions. This approach is different from the one adopted here in several respects. First, we assume that SCls are always generated in INFL. Second, cliticization is done via V-Raising to INFL in the phonological component; cf. Lobeck and Roberge (1985) and Roberge (in prep) for detailed analyses.

[2] The term Colloquial Fr. refers to the dialect spoken in Québec and in some regions of France.

[3] The Trentino data are taken from Brandi and Cordin (1981) and Jaeggli (1984).

[4] The particular analysis of this construction is not relevant for this paper.

[5] A fourth possible analysis could involve a violation of the Bijection Principle (Koopman and Sportiche 1981). However, for independent reasons, the position occupied by the SCl can be neither an A- nor an A'-position, hence the SCl is not a variable nor a binder; cf. Borer (1984).

[6] It is suggested in Roberge (in prep) that the appeal to the so-called dummy Case-markers is dictated by structural conditions which explain their use in OD but not in SD.

[7] Another potential problem for Jaeggli's analysis has to do with the Portuguese facts. In this language it is possible to extract the subject even though there is rich subject-verb agreement but no free inversion. It can therefore not be argued that subject extraction originates from a postverbal position in this null subject language; cf. Zubizarreta (1982).

[8] The term Clitic-chain but not the analysis or the definition is borrowed from Hurtado (1985).

[9] O. Jaeggli (p.c.) pointed out to me that his line of research follows a path similar to mine with respect to (31). The differences between the two approaches follow crucially from different conceptions of empty categories.

[10] The question of determining how a clitic doubled NP gets (or not) Case is a separate issue; cf. Roberge (in prep).

[11] Lexical pronouns are blocked by the Case Filter under a certain view of how Case is assigned in doubling constructions.

REFERENCES

Borer, H. (1984) **Parametric Syntax: Case Studies in Semitic and Romance Languages.** Foris, Dordrecht.

Bouchard, D. (1982) **On the Content of Empty Categories.** PhD dissertation, MIT.

Brandi, L. and P. Cordin (1981) On Clitics and Inflection in Some Italian Dialects. Ms. Scuola Normale Superiore, Pisa.

Chomsky, N. (1981) **Lectures on Government and Binding.** Foris, Dordrecht.

Chomsky, N. (1982) **Some Concepts and Consequences of the Theory of Government and Binding.** MIT Press, Cambridge.

Everett, D.L. (1985) Pírahã Clitic-doubling and the Parametrization of Nominal Clitics. Ms. MIT.

Hornstein, N. (1984) **Logic as Grammar.** MIT Press, Cambridge.

Hurtado, A. (1985) Clitic Chains. Ms. Simon Fraser University.

Jaeggli, O.A. (1982) **Topics in Romance Syntax.** Foris, Dordrecht.

Jaeggli, O.A. (1984) Subject Extraction and the Null Subject Parameter. **NELS** 14.132-153.

Kayne, R.S. (1984) **Connectedness and Binary Branching.** Foris, Dordrecht.

Koopman, H. and D. Sportiche (1981) Variables and the Bijection Principle. **The Linguistic Review** 2.139-160.

Lobeck, A. and Y. Roberge (1985) Coordination, Verb Raising and Subject Clitics. Paper presented at the LSA meeting, Seattle.

May, R. (1977) **The Grammar of Quantification.** Distributed by the IULC.

Rizzi, L. (1985) Null Objects in Italian and the Theory of pro. Ms. MIT.

Roberge, Y. (1986a) Subject Doubling, Free Inversion and Null Argument Languages. **Canadian Journal of Linguistics** (to appear).

Roberge, Y. (1986b) On Doubling and Null Argument Languages. **NELS** 16.

Roberge, Y. (in prep.) **The Syntactic Recoverability of Null Arguments.** PhD dissertation, UBC.

Safir, K.J. (1985) **Syntactic Chains.** Cambridge University Press, Cambridge.

Sportiche, D. (1983) **Structural Invariance and Symmetry in Syntax.** PhD dissertation, MIT.

Zubizarreta, M.L. (1982) Theoretical Implications of Subject Extraction in Portuguese. Ms. MIT.

Donca Steriade

Gemination and the Proto-Romance Syllable Shift

1. Introduction

The problem discussed in this paper is that of the spontaneous shifts in syllable structure commonly assumed to have taken place between prehistoric Latin and the contemporary Romance languages.

Complex onsets consisting of obstruent-liquid sequences must be posited at three stages in the history of Latin-Romance: throughout the entire recent history of each Romance language, in classical Latin and during at least one early stage of prehistoric Latin. However, despite the obvious identity between early Latin, classical Latin and contemporary Romance onset structure, it has been impossible so far to claim complete historical continuity between these stages. Two sound changes have been identified as the source of this difficulty: the Proto-Romance stress shift from antepenult to penult in words like **integra** and the Pre-Latin raising of reduced vowels in items like **con-secro** (from /con-sacr-o/). Both phenomena have been widely interpreted as indicating that at one or more intermediate stages in the history of Latin-Romance obstruent-liquid clusters were heterosyllabic. It is the purpose of this paper to show that obstruent-liquid clusters have been tautosyllabic throughout the entire history of Latin-Romance and that both the Proto-Romance accent shift and the Pre-Latin raising of **a** to **e** in **consecro** require alternative explanations. I begin by making explicit the historical view of Latin-Romance syllabification that must underlie the traditional interpretation of these phenomena.

2. Stages in Latin-Romance syllable structure: a reconstruction of the standard view

2.1. Metathesis and -lo-syncope in early Latin

The earliest Latin sound change differentiating between light and heavy syllables has been identified by A.-C. Juret (1921) as the metathesis of a consonantal sonorant with a following vowel. According to Juret, this sound change took place only if the sonorant was preceded by another segment **within the same syllable**: "Une explosive devient implosive dans la même syllabe; cette explosive est toujours **r**, **l** ou **n** précédée d'une autre consonne initiale de syllabe; lorsque **r**, **l**, **n** sont seules initiales de syllabe, elles se maintiennent après consonne comme après voyelle: **vulnerare** ne peut devenir **vulenrare**" (Juret 1921: 244). Some cases for which metathesis is clearly responsible include **tarpezita** for **trapezita** 'money-changer', **pulmo** for ***plumo** 'lung' (cf. Greek **pleumon**) **dulcis** for ***dlukui-** 'sweet' (cf. Greek **glukus**).

Juret used metathesis to argue that all intervocalic obstruent-sonorant sequences of Latin were tautosyllabic during this early stage. In his view ***sacro-lo-m** metathesized to **sacorlom**, which eventually became classical **sacellum** 'small shrine'.[1] Similarly **signo-lo-m** would have metathesized to **sigonlom**, later to become classical **sigillum** 'small mark'. Metathesis must then have operated at a stage when **sacrolom** and **signolom** were sylalbified **sa.cro.lom**, **si.gno.lom**.

A complete search of the diminutive formations in **-elo**, Juret's main source of evidence, leads to a conclusion that differs in detail but not in substance from Juret's reconstrution. We may divide the sequences preceding the **-elo** suffix in three classes: (a) VC_1 sequences as in **asin-elo-s** 'little donkey'; (b) VC_1C_2 sequences where the C_1C_2 cluster is of increasing sonority, as in **sacr-elo-s**; and (c) VC_1C_2 sequences where C_1C_2 is of decreasing sonority, as in **vern-elo-s**. One observes loss of the suffixal **e** in class (a) (**asinelos** → **asinlos**, eventually **asellus**); one observes an apparent reordering of **e** and the immediately preceding sonorant in class (b) **sacr-elo-s** → **sacerlos**, eventually **sacellus**); and one notes that no change takes place in the syllable structure of the last class:

(1)

a. VC-elo- diminutives

asin-us 'ass'; *asin-elo-s → *asinlos → asellus
amphor-a 'amphor'; *amphor-ela → *amporla → ampulla

b. VCC-elo- diminutives, where CC is of increasing sonority

*sacros 'sacred' → sacer; *sacr-elo-m → *sacrlom → sacellum
*agros 'field' → ager; *agr-elo-s → *agrlos → agellus
cerebrum 'brain'; *cerebr-elo-m → *cerebrlum → cerebellum
pugnus 'fist'; *pugn-elo-s → *pugnlos → pugillus
columna 'column'; *column-ela → *columnla → columella

c. VCC-elo- diminutives, where CC is of decreasing sonotiry

cornu 'horn'; corniculus (not *corn-elo-s → cornlos → *corellus)
verna 'slave born in the house'; vernulus (not *vern-elo-s → *verellus)
gland- 'acorn'; glandulae (not *glandlae, or *glallae)
cista 'box'; cistula, cistella (not *cistla or *cilla)
calc-s 'stone'; calculus (not *calclus)

The generalization appears to be that the vowel **e** in **-elo-** is elided after a single consonant or a cluster of increasing sonority (such as **cr**, **gr**, **br** or **gn**) but not after a cluster of decreasing sonority (such as **st**, **lc**, **rn**). It is safe to assume that this last class of sequences was heterosyllabic at this stage, as throughout the history of Indo-European languages. We may hypothesize then that the one context where e-Elision in diminutives could not take place was after a closed syllable, as in **ver.n-e.lo-s**. The natural extension of this hypothesis is that all clusters after which **e** could be deleted represent complex onsets. We are led then to postulate syllabic divisions such as **ce.re.bre.lom**, **sa.cro.lom** etc. and the rule of e-Elision below:

(2) **e-Elision**

$$e \rightarrow o \;/\; V\;[C.\overset{\top}{V}]$$

It is entirely possible that **e-Elision** is simply a restricted version of the general rule of syncope active in prehistoric Latin (cf. Vendryes 1907).

The **nl** and **rl** clusters created by elision (cf. intermediate ***asinlos**, ***sacerlos**) underwent a general rule of Latin, discussed below, and became geminate **ll**. Where elision gave rise to an unsyllabifiable triconsonantal cluster an equally general epenthesis rule inserted **e**: hence intermediate ***sacrlom** - with a syllabically stray **r** - became ***sacerlom**, later **sacellum**.[2]

Although this scenario for the derivation of **-elo** diminutives differs from Juret's, his essential idea is upheld: cluster such as **cr** have a different effect in the formation of the diminutives from heterosyllabic consonant sequences such as **lc**. Moreover, Juret's observations on the operation of metathesis proper appear well justified: metathesis always affects complex onsets (**trapezita** → **tarpezita**), never heterosyllabic C-sonorant clusters (such as **vul.ne.ra.re**, ***vulenrare**).

To sum up this portion of the evidence: **e-Elision** and the sporadic metathesis of a postconsonantal sonorant took place at a stage in the history of Latin when obstruent-liquid clusters, as perhaps all consonant clusters of increasing sonority, formed complex onsets. The fact that consonant clusters such as **gn** behave as onset sequences with respect to **e-Elision** is strongly reminiscent of the type of complex onsets reconstructible for Indo-European on the basis of Mycenaean, Sanskrit and other Latin evidence (cf. Steriade 1982 and forthcoming). For this reason, I assume that the syllabic divisions indicated by **e-Elision** represent the earliest stage of Latin syllabification.

2.2. Vowel reduction in early Latin

Short vowels in medial syllables underwent a number of related changes during the earliest documented stages of Latin. In open syllables all vowels raised to a high vowel whose backness/roundness matched that of neighboring consonants. In closed syllables **a** and **e** became **e**; o raised to **u**. Finally **i** and **u** ramained unchanged. The remarkable aspect of this development is that before obstruent-r clusters a short **a** raised to **e**, not **i**, as if contained in a closed syllable: **con-sacro** became **consecro**, rather than the expected ***consicro**. Since only this last aspect of vowel reduction is relevant for present purposes, I limit the examples to the normal development of medial short **a**.

(3) Vowel reduction: short medial **a**

a. before a single C
ago 'set in motion'; ex-ago → exigo 'drive out'
facio 'make'; con-facio → conficio 'fashion'
[y]acio 'throw'; con-[y]acio → con-[y]icio → con-icio 'throw together'
cado 'fall'; con-cado → concido 'collapse'

b. before two heterosyllable C's
arceo 'protect'; ex-arceo → exerceo 'train, exercise'
[y]acto 'fling'; con-[y]acto → con-[y]ecto 'fling together'
damno 'punish'; con-demno 'condemn'
factur 'made'; con-factus → confectus 'fashioned'

c. before obstruent-r
sacr- 'sacred'; con-sacro → consecro 'consecrate'
tag- 'touch'; in-tag-ro-m → integrom → integrum 'untouched, whole'
patro 'accomplish'; per-patro → perpetro 'perpetrate'
im-patro - impetro 'obtain by asking'

The pattern of raising in (3)c has been in general interpreted as evidence for the syllabification **con.sac.ro.** This hypothesis goes back at least to E. Hermann (1917 and 1923) and is adopted in much of the later work on Latin phonology: cf. Leumann 1977:83 and Safarewicz 1969.

Taken at face value, the facts of Raising before Obstruent-r clusters require that two drastic changes in the syllable structure of early Latin be posited. If the Raising that produced **consecro** reflects a later stage than that uncovered by Juret then the historical development must pass through the following steps: a prehistoric syllabic assignment **(con).sa.cro** which feeds e-Elision, a more recent assignment **(con).sac.ro** required for Raising, followed by the pre-classical and classical syllabic division **con.se.cro, sa.cro.** Neither the early shift from **sa.cro** to **sac.ro** nor the later shift back to **sa.cro** can be tied to any other developments in the phonological system of Latin.

2.3. (Pre)Classical Latin: stress and meter

Since about the 4th century BC, Latin stressed the penultimate syllable, if heavy, and otherwise the antepenultimate. Counting as heavy is any syllable containing a long vowel or diphthong and any closed syllable. Penultimate syllables containing a short vowel followed by an obstruent-liquid cluster (as in **vólucris, íntegra, lócuples, ténebra**) appear to be light: the accent falls on the antepenult in such cases.[3] The testimony of the accent is supported by the facts of metrical scansion, in particular by the systematic scansion of V -obstruent liquid -V sequences as $\breve{\sigma}\ \sigma$ in the comic poets Plautus and Terentius.

2.4. Vulgar Latin and Proto-Romance

In the transition from Latin to Romance, the location of the accent was generally preserved.[4] Thus the contrast between antepenult stressed Classical **púlic-em** 'flea-ACC.sg.' and penult stressed **félīc-em** 'happy-ACC.sg' is still preserved in Romanian **púrice** vs. **feríce.** In particular antepenult stress in words of the form [...V́CV̆CV(C)] is observed unchanged throughout the entire Romance family: **dómina** 'lady', **dómine** 'lord-VOC' become Rom. **dóamna, dóamne** in Romanian, It. **dónna** (from **dómna** with medial syncope of the unaccented vowel); **nébula** 'cloud, darkness' becomes Rom. **négura; basílica** 'church' becomes Rom. **bisérica; iúdicem** becomes Italian **guídice**, Rom. **júdece; iúvenem** becomes Ital. **gióvine**, Rom. **júne** (from ***júβene**, see below); **léporem** becomes Rom. **iépure**, Fr. **lièvre** (with medial syncope of unaccented **o**).

The Latin accent shifts rightwards in Proto-Romance when the penult is separated from the ultima by an obstruent-liquid cluster: **íntegrum** → **intégrum**, hence Rom. **intégru** (with metathesis of **r**), It. **intéro** (with spirantization and loss of **g**) and Fr. **entiér** (also with loss of **g**). It also shifts rightwards when the antepenult and penult are in hiatus: **filíolus** → **filiólus**, hence Fr. **filleúl**; **mulíerem** → **muliérem**, hence Sp. **mujér**, Rom. **muiére**.

Let us first eliminate from the discussion the rightwards stress shift in **mulíerem** → **muliérem**. This phenomenon is simultaneous with the formation of Proto-Romance postconsonantal glides: the Latin hiatus between a non-low vowel and a following vowel (**mu.lí.e.rem**, **há.be.at** etc.) is replaced by the Proto-Romance syllabification **mu.lyé.rem**, **há.byat**. In this context, the accentual change is easily explained along the lines suggested by the analyses of similar stress shift cases given in Al-Mozainy et al. (1985), Halle and Vergnaud (forthcoming) and especially Harris (1986): stress shifts within a binary foot from the desyllabified foot head onto the remaining foot syllable. Thus the foot structure of Classical **mulíerem** must be **mu. li.e (rem)**, with an extrametrical final syllable (cf. Hayes 1982). Primary stress falls on the rightmost foot and, within it, on the leftmost syllable **li**. When the Proto-Romance glide formation eliminates **li** as a syllable, the remaining vowel **e** inherits the accentual properties of the entire foot. The only sound change that must be recorded is then the glide formation: the direction of accent shift triggered by glide formation is fully predicted by a stress theory that recognizes metrical constituent structure.

We may turn now to the stress shift from antepenult to penult in words of the **íntegrum** class. Suppose that the Proto-Romance assignment of obstruent-liquid clusters is heterosyllabic: **in.teg.ra** rather than **in.te.gra**. Then the stress predicted by the Latin stress rule should indeed be penultimate: **in.tég.ra**. This is, with minor variations, the explanation offered by every scholar that has written on this subject. This explanation commits one to two more spontaneous syllable shifts: the shift from classical **ín.te.gra** to Romance **in.tég.ra** and the shift back to **in.té.gra**, needed to account for the syllabification facts of contemporary Romance (on which see below).

The clearly problematic aspect of this explanation is that one fact - the shift in stress - is given as justification for not one but two changes in the successive grammars of Romance, the second change being needed in order to cancel entirely the effects of the first. Various improbable improvements on this scenario have been proposed, the most widely quoted being Hermann's (1917, 1923) hypothesis[5] that spoken Latin had never gone through the **in.te.grum**, **sa.crum** stage at which obstruent-liquid clusters are complex onsets. Rather, Hermann suggests that spoken Latin - as against literary, hochsprächlich Latin - would have preserved the heterosyllabic assignment of obstruent-liquid clusters, from the prehistoric stage reconstructible from the effects of Raising in **consecrare** into the Proto-Romance stage reconstructible from the penult stress in **intégra**. One of the odd corollaries of this idea is that the stress pattern unanimously reported by Latin writers and grammarians is a literary fiction. Another is that the language of Plautus, where obstruent-liquid clusters are systematically tautosyllabic, is hochsprächlich in its syllabic divisions and colloquial in everything else. In addition, we are still left in the dark as to how the Romance languages, which on Hermann's hypothesis must have inherited the closed syllabification of spoken Latin, succeeded in acquiring independently of each other the very same onset structure as that of "literary" and Plautinian Latin.

This last point is worth developing in some detail. As in Plautinian Latin, French (Selkirk 1978), Italian (Vogel 1977), Portuguese (Reighard 1982), Spanish (Harris 1983) and Romanian (Steriade 1984) syllabify obstruent-liquid sequences as onsets. It is hard to imagine that the syllabification of contemporary Romance languages resulted from independent innovations; although many unrelated languages allow a certain subset of obstruent-sonorant clusters as complex onsets, the details of such onset inventories tend to vary substantially from language to language.[6] In contrast, Latin and the Romance languages have onset inventories that are identical down to the details: **tl**, **dl** clusters are heterosyllabic in Latin, as well as Romanian and French.[7] Moreover, it is possible to show that the contemporary tautosyllabic assignment of obstruent-liquid clusters was already present at the earliest stages in the history of individual Romance languages. For example, the 10th century Tuscan diphthongization of stressed mid vowels in open syllables takes place in **pé.de → piéde** 'foot' as well as in **pé.tra → piétra** 'stone', but not in **fés.ta** 'holiday' (Rohlfs 1966:25). The earliest French diphthongization (5th-6th century, according to Fouché 1969:223) affects the open syllables of **ad.ré.tro → arrière** 'behind', **fé.bre → fièvre** 'fever', but not the closed syllables of **fes.ta →** O.Fr. **feste, *gen.ru → gendre** 'son-in-law' (Fouche 1969:223). The most natural conclusion to be drawn from these considerations is then that the onset structure of contemporary Romance languages cannot be the result of independent innovations: the Romance language are descended from a dialect in which words like **integra**, however stressed, contained an open penult. More evidence for this view will be presented below.

3. Guidelines for reconstruction

The primary data usable in the reconstruction of Romance phonology are sequences of segments preserved in the Romance languages. The syllabification of these sequences cannot be directly observed, as no metrical texts can be confidently ascribed to the variety of spoken Late Latin that we call Proto-Romance. Proto-Romance syllabification must then be inferred: the ideas guiding this process are enumerated below.

3.1. Onsets are maximized

One observes frequently that an intervocalic consonant is always syllabified as onset to the second syllable. More significant for our purposes is the observation that a consonant sequence that can be syllabified as an onset in a given language, must be so syllabified.

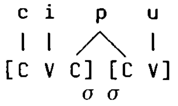
3.2. Surface syllabification is exhaustive

A typological study of consonant cluster simplifications shows that, given independent criteria for what is a possible onset and what is a possible coda in a given language, only those clusters surface that can be exhaustively parsed into coda-onset sequences. The most common type of cluster simplification is the loss of C_2 in a medial sequence $VC_1C_2C_3V$ where C_1C_2 cannot be a coda and C_2C_3 cannot be an onset.

From the two generalizations formulated above two others follow: if in a medial sequence VC_1C_2V the consonant cluster is heterosyllabic ($VC_1.C_2V$), then C_1C_2 can surface in postconsonantal position ($VC_3C_1C_2V$) only if the cluster C_3C_1 is a possible coda in the language. Conversely, if C_3C_1 is known to be an impossible coda, then the existence of surface sequences $VC_3C_1C_2V$ indicates that C_1C_2 is a possible onset.

3.3. Rules that may not affect geminate segments refer to syllabic constituents or syllable boundaries

A set of related Romance sound changes voice and spirantize postvocalic obstruents when followed by vowel or liquid: e.g. **ripa** → Fr. **rive**, **capra** → Fr. **chèvre**. The general context of these rules can be formulated either as [-cons]_ [+son, +cont] or, assuming that obstruent-liquid clusters are onsets, as [-cons] $[_{\sigma}$_(after a vocoid and at the beginning of a syllable). The choice of formulations is not free: is we adopt the latter, we predict that these sound changes will never affect geminate segments, since in a form like **cippu**,



the postvocalic segment **p** is simultaneously at the beginning of one syllable and at the end of the other, and thus does not satisfy unambiguously the structural description of voicing/spirantization (cf. Schein and Steriade 1986). On the other hand, if we adopt [-cons]__ [+son, +cont] as the context of voicing/spirantization, then we predict that geminate obstruents will voice and spirantize, since the **segment p** in **cippu** is indeed preceded by a [-consonantal] segment and followed by a continuant sonorant. Conversely, if we know that voicing/spirantization fails to affect geminates, we would be naturally led to opt for a formulation of these rules that will predict their inapplicability to geminates: we would choose the context $[\text{-cons}]\ [_{\sigma}\text{—}]$ and the auxiliary assumption that obstruent-liquid clusters may occur in syllable-initial position.

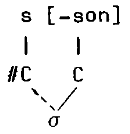
3.4. The boundaries of cyclic domains may correspond to syllabic boundaries

Cyclic assignment of syllable structure frequently results in apparent contrasts of syllabic division. Thus the cluster **br** is tautosyllabic in Latin **labrum** 'lip', as everywhere stem-internally, but heterosyllabic in **ab-rogo** 'to annull' and in every case where a prefix or compound boundary separates its two members. The explanation for the contrast may be this: **labrum** contains a single cyclic domain, within which the principle of onset maximization requires that **br** be assigned in toto to the second syllable; but **ab-rogo** contains two independently syllabified domains, **ab** and **rogo**. The concatenation of fully syllabified **ab** and **rogo** yields, predictably, **ab.ro.go**. This observation is of great significance to the present discussion; if the cluster **br** is heterosyllabic in some context then it is expected to behave phonologically like the heterosyllabic **br** cluster of **ab-rogo**.

3.5. Heterosyllabic initials and stray adjunction

Latin word-initial clusters are not limited to onsets. An **s** can precede any onset whose initial is a stop: clusters like **st** (stō 'I sit', **sk** (**scelus** 'crime'), **str** (**struo** 'put in order'), **spr** (**spretiō** 'contempt'), **spl** (**splendeō** 'shine') occur initially even though the same sequences are heterosyllabic in intervocalic position (**es.tis** 'you are', **as.tra** 'stars', **as.pra** 'rough-feminine', **re-s.plen.de.ō** 'to glitter back'. An analysis of this pattern was proposed in Steriade (1982): a rule of stray adjunction provides for the syllabic incorporation of an initial unsyllabified **s**:

(4) Latin Stray Adjunction



Since this rule follows the general process of word-internal syllabification it will allow for the creation of clusters like **str**, where **s** is adjoined to an already complex onset. For the same ordering reason, the syllabic incorporation of initial s will be bled by coda-incorporation in sequences like **re-s.ti.tu.ō**, **re-s.plen.de.ō.** Note that regular complex onsets in the same position do not give rise to the same syllabification pattern: **re.-fre.nō**, **re.-pli.cō.**

S is the only segment to which the Latin Stray Adjunction applies. It follows then, from (4) and from the principle of exhaustive syllabification, that no clusters other than (s)[-son])([+son,+cont]) may occur word-initially in Latin. Indeed, a variety of other clusters, some native to Latin and some belonging to the loan vocabulary, are simplified in this position by losing their first member, the stray consonant for which no adjunction provision exists:

(5)

#tl > l *tlātos > lātus 'carry-participle'

#dl > (ll) > l *dlongos > longus 'long'

#gn > ŋn > n gnixus 'lean-pple' > nixus
gnō- as in i-gnōtus 'unknown', co-gnōscō 'learn' but nōscō 'know', nōmen 'information, name', nōtus 'known'

#kt > t *ktunika > tunica 'shirt' cf. Hebr. ktionet gk. Ktesiphon > Tesifon

#pt > t gk. ptisane 'barley-gruel' > tisana
kg. Ptolomaid- > Tolomaidi

The Latin treatment of heterosyllabic initials is encountered, with minor variations, in Greek, Sanskrit, in certain Arabic dialects, in Klamath and in many modern Indo-European languages. Abstracting away from the language specific details of stray adjunction rules, one can conclude generally that a cluster C_1C_2 that surfaces as heterosyllabic in intervocalic position can occur word initially only if a stray adjunction rule exists that provides for the incorporation of C_1 in an initial syllable beginning with C_2. This generalization will also play a role the reconstrution of Proto-Romance syllabic divisions.

4. Syllable-based generalizations in Proto-Romance

Equipped with these guidelines, we may now proceed to determine the syllabic status of Proto-Romance obstruent-liquid clusters.

If, at any point during the history of Proto-Romance, obstruent-liquid clusters had ceased to be possible onsets then we would expect a number of accompanying changes:

(a) Neutralization of the contrast between heterosyllabic **br** (**ab-rogo**) and tautosyllabic **br** (**la.bra**): both clusters should behave like the former with respect to syllable-based rules.

(b) Simplification of certain post-consonantal clusters: a sequence $VC_1C_2C_3V$, where C_2 is an obstruent and C_3 is a liquid should surface only if C_1C_2 is a possible coda cluster. For instance, if the coda clusters of both Latin and Romance are of decreasing sonority, one expects that $VC_1C_2C_3V$ should lose C_2 if both C_1 and C_2 are obstruents.

(c) Simplification of certain initial clusters: if obstruent-liquid clusters cease to be possible onsets one normally expects that such sequences would be preserved only in contexts where their first member can become coda to a preceding syllable. Initially, such clusters should lose their first member.

In the remainder of this section I will present the Romance evidence that disconfirms each one of these predictions.

One should bear in mind, however, that the predictions hold only if the single relevant difference between the phonological system of Latin and that of Proto-Romance is the change in onset structure discussed here and if every thing else is held constant. Each prediction can be voided by adding to the hypothesis of heterosyllabic obstruent liquid sequences various other auxiliary hypotheses. For example, the patently false prediction in (c) - that word-initial obstruent liquid clusters should lose their first member - can be neutralized by postulating a Proto-Romance stray adjunction rule that would allow the syllabic incorporation of an initial obstruent before a continuant sonorant. It is no doubt possible to resort to similar moves in the face of evidence falsifying other predictions of the heterosyllabic hypothesis. The very least I hope to show, then, is that not a single prediction of this hypothesis can be confirmed. Some of the data discussed below make however for a stronger argument and indicate that a better understanding of Romance phonology can be gained if we assume that the Romance onset clusters are the same as those of Classical Latin.

4.1. Regressive Gemination

Consider first the contrast between **ab.ro.go** and **la.bra**. There exists a phonological rule in Latin that is sensitive to this contrast and a Romance continuation of this rule. In Latin, certain heterosyllabic consonant clusters are subject to a regressive gemination rule. To identify the class of clusters subject to gemination, I distinguish below three types of heterosyllabic clusters: (a) clusters of decreasing sonority; (b) clusters of increasing or equal sonority of the form [-coronal, αsonorant, αnasal] [+coronal, αsonorant, nasal]; (c) all other heterosyllabic clusters. One observes that classes (a) and (b) are preserved, while class (c) undergoes regressive gemination.

(6) Heterosyllabic clusters in Latin
(a) decreasing sonority
(i) liquid C: arvus 'field', arma 'weapons', rursus 'backwards', arbos 'tree', artis 'art-Genetive', pelvis 'pelvis', vulnus 'wound', alter 'other',
(ii) nasal C: inter 'between', infra 'under'; imperō 'command', unguō 'annoint', con-temnō 'despise',
(iii) sC: aster 'star', aspra 'rough', dis-cedō 'separate', dis-par 'unlike', dis-tineō 'keep apart'.

(b) [-cor, αson, αnas] [+cor, son, nas]
(i) [-son] clusters: factum 'deed', aptus 'able', re[ks]ī 'I reigned', ipse 'himself', ob-secrō 'beseech', ob-tineō 'hold',
(ii) [+son] clusters: amnis 'river', sollemne 'solemn', li[ŋ]num 'wood', si[ŋ]num 'sign'

(c) other clusters
(i) C-y
per-[y]erō → per.[y]e.rō → pe[yy]erō spelled <peiierō>, 'perjure oneself'
con-[y]iciō → con.[y]i.ci.ō → co[yy]iciō <coiiciō> 'throw'[9]
dis-[y]ungō → dis.[y]un.gō → di[yy]ungō <diiungō> 'separate'

(ii) C-r
sub-ripiō → sub.ri.pi.ō → sur.ri.pi.ō 'steal'
BUT la.bra 'lips'
ad-ripiō → ad.ri.pi.ō → ar.ri.pi.ō 'snatch'
BUT la.drō 'thief'
in-ruō → in.ru.ō → irruō 'rush in'

(iii) C-l
ad-loquor → ad.lo.quor → al.lo.quor 'adress, speak to'
lapid-elo-s → la.pid.lus → lapillus 'pebble'
per-legere → per.le.ge.re → pellegere 'scan'

ager-elo-s → a.ger.lus → agellus 'little field'
coron-el-a → co.ron.la → corolla 'little crown'
in-lūstris → in.lūs.tris → illūstris 'famous'

(iv) C-nasal
ad-nuō → ad.nu.ō → annuō 'nod in assent'
merced-nārius → mer.ced.nā.ri.us → mercennārius 'hired soldier'
sub-moveō → sub.mo.ve.ō → sum.mo.ve.ō 'remove'
ad-mittō → ad.mit.tō → am.mit.tō 'send to'
glūb-ma → glūb.ma → glūm.ma → glū.ma 'peel (of fruit)'[10]
pesna (Old Latin) → pes.na → penna 'feather'
dis-minuō → dis.mi.nu.ō → dimminuō 'dash to pieces'
in-mineō → in.mi.ne.ō → immineō 'threaten'

(v) C-obstruent
sitcus → sit.cus → siccus 'dry' (cf. sitis 'thirst')
sub-gerō → sub.ge.rō → suggerō 'provide'
ob-ferō → ob.fe.rō → offerō 'offer'
ob-cupō → ob.cu.pō → occupō 'occupy'
dis-ferō → dis.fe.rō → differō 'differ'
ec-ferō → ec.fe.rō → efferō 'carry out'
ad-plico → ad.pli.co → applico 'attach to'
ad-sideō → ad.si.de.ō → assideō 'sit by'
con-cut-sī → con.cut.sī → concussī 'I struck'
quid-[kw]am → quid.[kw]am → qui[kkw]am, spelled <quicquam>[11]

As the examples in (6a) indicate, sequences of decreasing sonority are not subject to regressive gemination and remain, in general, intact.[12] When the sonority profile of the consonant sequence is not decreasing, two treatments are observed: the [-coronal][+coronal] clusters of identical obstruency and nasality are preserved while all other clusters geminate. Gemination takes place both in sequences of level sonority (sf → ff; cp → pp, nm → mm, rl → ll) and in sequences of increasing sonority (ts → ss, sm → mm, dm → mm, bm → mm, br → rr, dr → rr, dl → ll).

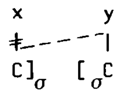
Before formulating the rule, a number of details must be mentioned. Only some underlying instances of **gm** yield predicted **mm**: underlying **cm** becomes **gm** and, unaccountably, fails to undergo gemination. Thus **flag-ma** 'flame' becomes **flamma**, Greek **sigma** becomes **simma** (Schuchardt 1866:146) but **sec-mentum** becomes **segmentum**, not ***semmentum**. A possibly related fact is that syllable final g becomes **w** before m in Late Latin: **fragmentum** is often written **fraumentum**. If the date of the g → w change is in fact earlier, then perhaps this rules bleeds Regressive Gemination by creating **wm**, a cluster of decreasing sonority.

Across a prefix boundary, gemination seems wide-spread and perhaps obligatory when it involves two obstruents or two sonorants: **dis-ferō** → **differō**; **in-ruō** → **irruō**. On the other hand, when prefixation consists of obstruent-sonorant, gemination is clearly optional: both **arrideō** and **ad-rideō**, both **ad-moveō** and **ammoveō** are encountered. Finally, no classical Latin texts record the possibility of gemination with **b.l** sequences: only **ob-loquor** occurs, never ***olloquor**.

In some cases, Regressive Gemination is preempted by other processes: **ml** and **mr** undergo epenthesis and become **mbr**, **mbl** respectively; **bn**, **bm**, **gn** assimilate in nasality and yield **mn**, **mm** and **ŋn**. Manuscript spellings like **ab-negō**, which seem to violate both the **bn** → **mn** rule and Regressive Gemination, are contradicted by inscriptions, where the expected **amnegō** occurs (cf. Niedermann 1953:135).

The irregularities of Regressive Gemination should not obscure its importance for Latin phonotactics: together with the syllabification process, this rule explains the very limited inventory of Latin medial clusters.

(7) Latin Regressive Gemination



unless (a) y is of lower sonority than x or
(b) xy is [-cor,αson,αnas][+cor,αson,αnas]

optional in [-son][+son] sequences; inapplicable to b.l.

For our purposes, the relevant fact is that regressive gemination distinguishes between tautosyllabic obstruent-liquid clusters and heterosyllabic ones: recall **sub.ripiō** → **surripiō** vs. intact **la.bra**.

In the transition from Latin to Romance, Regressive Gemination changed along predictable lines, by eliminating some of its conditions. It began to apply regularly to heterosyllabic **bl** sequences and to all [-son][+son] sequences resulting from prefixation. It was extended to all clusters of increasing or level sonority, including to the [-cor][+cor] clusters. The new rule applied obligatorily across a boundary (in derived environments, as a cyclic rule) and optionally stem-internally (postcyclically). Later, individual Romance languages like Italian made the rule obligatory in all contexts while most others lost it. However, and this is significant, Romance Regressive Gemination never applied to stem-internal clusters of obstruent-liquid: the contrast between geminating **b.r** in **sub-ripio** and non-geminating **br** in **labra** is preserved in Romance.

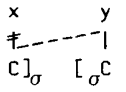
(8)

cyclic: ob-vius → *ovvius → It. ovvio
ad-versarius → *avversariu → It. avversario, Old Fr. aversier
co-[ŋ]nosco → *connosco → Rom. cunosc, Fr. connais
ad-retro → *arretro → Fr. arrière
ad-satis → *assatis → Fr. assez, Old. Sp. asaz
ad-mutio → *ammutio → Rom. amutesc
sub-levio → *sullevio → Fr. soulager, Sp. soliviar
sub leone → *sulleone → It. sollione
sub-ridere → *suridere → It. sorridere, Fr. sourir, Rom. suride
*ab-runco → *arrunco → Rom. arunc
sub-terrare → *sutterare → It. sotterare, Prov. Sp. soterrar
ab-solvere → It. assolvere, Old Fr. assoldre

postcyclic: spatula → *spatla → Prov. spatla
*spalla → It. spalla, Fr. epaule
planula → *planla → It. pialla
scamnum → *scannum → Sp. escano and
*scamnum → Rom. scaun
li[ŋ]num → *linnum → Sard. linnu
*liŋnum → It. legno, Rom. lemn
staŋnare → *staŋnare → Sard. stanzare
fraksinus → *frassinu → Rom. frasin
co[ks]a → *coksa → Rom. coapsa
lactuca → lattuca (301 AD) → It. lattuga,
Cat. llatuga
→ lactuca → Fr. laitue
frigdor → *friddor → Sp. Cat. fredor and
*frigdor → Old. Port., Prov.
freidor
amygdala → amiddula (Appendix Probi: late
Latin)
*aptillare → *attilare → Fr. atteler
aptare → *attare → Sp., Prt. Cat. atar
scripsit → *scrissi → Rom. scrise
septem → *septe → Rom. şapte
→ *sette → It. sette

Some of the more important sound changes operating above are: the simplification of geminates in all Romance dialects other than Italian (***frassinu → Rom. frasin**); the Romanian change of a velar stop into a labial before t, n, s (**coksa → coapsa**); the weakening of a syllable final velar to i in Western Romance (**lactem → Fr. lait**).

(9) Romance Regressive Gemination



unless y is of lower sonority than x

cyclic : obligatory
postcyclic : optional

As a result of the Romance Regressive Gemination rule, no heterosyllabic clusters of equal or increasing sonority are preserved across a prefixal boundary in the Romance languages; and only very few such clusters (most of the class [-cor,αson, αnas][+cor,αson, αnas] are reconstructible for Proto-Romance stem-internally. Weinrich (1958-230) reaches essentially the same conclusion: "es gibt keine Konsonantengruppen mehr [in Proto-Romance] mit einem Verschlusslaut als erstes Element, ausser wenn das zweite Element **r** oder **l** ist (Muta cum Liquida)". Conversely, adds Weinrich, the only possible biconsonantal clusters in late spoken Latin are, aside from obstruent-liquid sequences, the clusters of decreasing sonority: liquid-C (**cardo, valde**), s-stop (**vespa, vestis, piscis**), nasal-obstruent (**campus, inde, longus**).

I sketch below the treatment of stem-internal obstruent-liquid clusters in Romance. The generalization that emerges is that these fail to undergo the Romance Regressive Gemination. Two conclusions will follow: the Romance rule, like its Latin counterpart, applies only to heterosyllabic clusters; and that stem-internal obstruent-liquid clusters are tautosyllabic.

4.2. Romance postvocalic onsets

Obstruent-l clusters were preserved occasionally in French (cf. ***aboclu** → **aveugle** 'blind'; ***duplu** → **double**) and systematically in Istrian (e.g. **auricla** → **urecla** 'ear'). Elsewhere l became y in such clusters and the resulting sequence underwent further, less general, changes: ***tribla** 'threshing sledge' → ***tribya** → Ital. **trebbia**; **auricla** → ***aurikya** → Rom. **urekye** (spelled <ureche>).[13]

The Latin obstruent-r clusters are generally preserved in Romanian, Spanish, and Southern Italian:

(10) nigrum → Rom. negru 'black'
febbruariu → It. febbraio 'february' (COMPARE subridere → sorridere)
*quadrus → Rom. codru 'large forest' (COMPARE adrunco → *arrunc, arunc)
capra → Rom., Ital. capra 'goat'
lacrima → Rom. lacrima, It. lagrima, Sp. lagrema 'tear'
petra → Rom. piatra, It. pietra, Sp. piedra 'stone'

In the Central Romance dialects (French, Provençal, Northern Italian and Tuscan) voiceless stops are voiced and spirantized before tautosyllabic **r**. At a relatively early stage the voiced spirant becomes a glide in preconsonantal position. The rules involved seem to be ordered differently depending on the clusters and on the individual dialects:

(11) integrum → *inteγru → *inteiru (Fr. entier)
acrum →* agrum (Fr. aigre): Voicing after
Spirantization)
patrem → *padrem → *paðre (Prov. paire):
Spirantization after Voicing

The rules of Voicing and Spirantization illustrated above applied not only to the postvocalic member of an obstruent-liquid cluster but also to intervocalic obstruents. If we assume that all stem-internal obstruent-liquid clusters are tautosyllabic in Romance, then the disjunction of [-cons]_V and [-cons]_Liquid contexts reduces to the environment **postvocalic and syllable initial**. Most Romance dialects had rules of voicing and spirantization in this context: the only variable factor is the class of obstruents subject to such rules. In Northern Italian the rule applied to all obstruents[14]. In Romanian, one of the more conservative dialects in this respect, spirantization is limited to a postvocalic, syllable-initial b, which becomes w. In intervocalic position this w is eliminated entirely: **libertare → *liwertare → *liertare →iertare** 'forgiving'. Before a liquid, however, w becomes u, a detail predictable from the fact that the intermediate **wr**, **wl** clusters must have been heterosyllabic in Proto-Romanian: **fabru → *faw.ru → fa.ur** 'artisan'; **stablum → *staw.lum → staul** 'stable', **cribu → criw.ru → ciur** 'sieve'. In contrast with the regular spirantization of the postvocalic onset **b**, Romanian coda **b** is occasionally preserved: **subtilis → subţire** 'thin', **sub.ter → sub(t)** 'under', in precisely those contexts where Romance Regressive Gemination had been optional. The relevant sound changes are outlined below:

(12) Romanian Spirantization

a. b → w/ [-cons] $[_\sigma$ —
b. w → Ø/ [-cons] $[_\sigma$ —

As noted above in (3.3), the context of Romance voicing and spirantization could in principle be restated in terms independent of syllabic divisions: [-cons]_ [+son, +cont]. The argument for the syllable-dependent context given in (12) is provided by the failure of geminate obstruents to undergo voicing or spirantization: contrast French **rive** (← **ri.pa**), **chèvre** (**ca.pra**) with **cep** (← **cip.pu**) or Romanian **bere** (← ***biwere** ← **bi.be.re**) with **gheb** (← **glib.bu**). In conjunction with a theory of geminate blockage such as the one presented in Schein and Steriade (1986), the syllable-dependent formulation of voicing and spirantization explains these contrasts. The alternative does not.

To sum up this part of the argument: Regressive Gemination (9), a streamlined version of the Latin rule (7), must be reconstructed for Proto-Romance, since its effects are regularly reflected in each Romance language. The Romance rule did not affect stem internal obstruent-liquid clusters, an indication that these sequences continued to be tautosyllabic. The early Romance sound changes that voice and spirantize postvocalic onsets affect the first member of obstruent-liquid clusters, demonstrating again their tautosyllabic status.[15]

4.2. Cluster Simplification

Latin coda clusters were of two types: nasal-obstruent (**emp.sī**, **iunc.tus**) and liquid-obstruent (**ar[ks]** 'citadel', **cal[ks]** 'heel'). Obstruent-s codas are attested only word-finally[16].

In the transition from Latin to Romance, codas became restricted to one member. First, already in Latin, liquid-obstruent codas simplified word-medially by losing their last member: **farc-tus** 'stuffed' became **fartus**, **ulc-tus**, **ulc-sī** 'avenge' became **ultus**, **ulsī**. At a later stage, nasal-obstruent codas lost their second member: **unc-tus**, **unc-sī** 'anoint' became **untus**, **unsī**, **temptō** 'to feel by touching' became **tentō**. False spellings like **lancterna** for **lanterna** (itself resulting from the simplification of Greek **lampter**), attested already in the 2nd century AD (Väänänen 1967:64), indicate a relatively early date for the complete elimination of complex codas from the system.

Since Proto-Romance had inherited a syllable type that disallows coda clusters, the analysis of any triconsonantal consonant cluster is disambiguated: $VC_1C_2C_3V$ could surface only if C_2C_3 was a possible onset. If complex onsets are eliminated from Proto-Romance, the triconsonantal clusters inherited from Latin should lose their middle member in all contexts. But if the onset structure of Proto-Romance is identical to that of Latin, then such clusters should surface if C_2 is an obstruent and C_3 is a liquid. The latter prediction turns out to be the correct one: clusters like **ltr** (**altra** ← Lat. **altera** 'other-feminine'), **cultrum** 'knife'), **scl** (masclus ← Lat. **masculus** 'male-dimin.'), **mbr** (**umbra** 'shade'), **mps** (**amplus** 'wide'), **ngl** (**ungla** ← Lat. **ungula** 'nail') are preserved in the individual Romance languages and suffer only dialect-specific sound changes, never the loss of the medial C_2.[17] The inherited triconsonant clusters were preserved because the complex codas had already been simplified in Late Latin and because no change had occurred in the structure of the onset.

Moreover, consonant clusters newly created within the early stages of the history of individual Romance languages were preserved under the same conditions: if they could be parsed into a single coda consonant and a Latin-style onset. In Provençal, for instance, one observes the contrast in the treatment of two syncope outputs: ***pertga** from **pertiga** (Lat. **pertica** 'pole, rod') simplifies to **perga** whereas **martror** from **martirorum** 'martir-Gen. pl' is preserved intact. Similarly **domestgar** from **domestigar** 'game' simplifies to **domesgar** whereas **anglu** from **angulu** is preserved (cf. Pellegrini 1960). The same observation can be abundantly illustrated with data from the early history of French (cf. Fouché (1966: 825-851)) and Romanian (cf. Nandriş (1963: 284-285)): syncope rules creating triconsonantal clusters are followed by the loss of C_2 in $C_1C_2C_3$ only when C_2C_3 is not a possible Latin onset.

4.3. Initial clusters

I summarize here the main facts about the development of Latin initial clusters in Romance and their implications for the reconstruction of Romance syllabic divisions.

The Romance languages maintain in general the Latin word-initial obstruent-liquid clusters. For the reconstruction of the Proto-Romance situation, the relevant generalization is that each type of Latin obstruent-liquid onset is reconstructible at the beginning of the word: **placere** 'please' → Rom. **placere**, Fr. **plaisir**; **frenu** 'rein' → Rom. **frîn**, It. **freno**, Fr. **frein**; **florem** 'flower' → Rom. **floare**, Fr. **fleur**; **blasfemiare** 'curse' → Rom. **blestema**, Sp. **blasma**, Fr. **blamer**; **breve** 'short' → It. **breve**, Fr. **bref**; **trabem** 'wood beam' Rom. **treaba**, It. **trave**, O.Fr. **tref**; **draconem** 'dragon' → Rom. **drac**, Fr. **dragon**, Cat. **drag**; **clavem** 'key' → Fr. **clef**, Rhet. **klaf**, Prov. **claus**; **glacies** 'ice' → Fr. **glace**, Prov. **glatz**; **crucem** 'cross' → Rom. **cruce**, It. **croce**, Fr. **croix**; **grana** 'grain' → Rom. **grîne**, It. **grana**, Fr. **grain**. The preservation of these clusters in initial position is predicted straightforwardly if their tautosyllabic status was transmitted from Latin to Romance.

Significantly, the Latin heterosyllabic initials were preserved less uniformly in Romance. Already in Late Latin initial s-obstruent clusters triggered an optional prothesis rule: inscriptions and late texts provide forms like **ispopondit** (Class. **spopondit**) 'he pledged', **espopondiderunt** (Class. **spoponderunt**), **iscripta** (Class. **scripta**) 'written). The prothesis of a short **i** or **e** was not limited to s-obstruent initials but applied generally to all heterosyllabic initials: **izmaragdus** (Greek **smaragdos**), **Ismurna** (Greek Smyrna), **eksenium** (Greek **ksenion**). I assume that the rule was postcyclic, since its cyclic application would produce intermediate unattested sequences in cases like **re-stituō** ***re-istituō**: a postcyclic prothesis rule will permit such forms to syllabify as **res.ti.tu.o** before it comes into effect.

(13) Romance Prothesis 1 (postcyclic, optional)

$$\emptyset \rightarrow V \,/\, \#\_\_\ C' \qquad C' = \text{unsyllabified } C$$

Because Romance Prothesis was optional, the attested Late Latin treatment of heterosyllabic initials shows three possible variants, tabulated below:

(14)	With Prothesis	Without Prothesis
rule (4) applic.	iscripta, espopondiderunt	scripta, spopondiderunt
rule (4) inapplic.	ismurna, eksenium	ptisane → tisana psalterium → salterium

Table (4) shows that where prothesis fails to apply, the Latin rule (4) continues to incorporate the initial stray **s**; but that heterosyllabic initials to which neither prothesis nor rule (4) has applied - the latter failing predictably before sequences other than s-[-sonorant] - simply lose their initial unsyllabified member, as comparable clusters had in Classical and Preclassical Latin. This, of course, is the direct consequence of the principle of exhaustive syllabification discussed earlier.

In this context, the failure of word-initial obstruent-liquid clusters as in **placere** or **crucem** to trigger either prothesis or cluster simplification becomes significant. On the hypothesis that these continued to be tautosyllabic in Late Latin, this is the predicted outcome: both prothesis and cluster simplification require an initial stray consonant in the postcyclic stage of the derivation, whereas obstruent-liquid initials have already been syllabified as onsets on the initial cycle. The idea that **obstruent-liquid** clusters were heterosyllabic in Romance makes, in its most natural form, the prediction that **placere**, **crucem** should similarly have varied between ***iplacere**, ***icrucem** and ***lacere**, ***rucem**.

The reflexes of the heterosyllabic initials in different Romance languages vary depending on the fate of rule (13). In Western Romance, where the rule became obligatory everywhere word-initially, heterosyllabic initials are preserved uniformly with a prothetic **e**: **sponsu** 'spouse' → Old Fr. **espos**, **schola** 'school' → Old Fr. **escole**, **smaragdu** → Old Fr. **esmeralde**. Fouché (1967: 703) mentions the fact that the Old French **pt**, **pn**, **ps**, **ks**, **hl**, **hr** heterosyllabic initials failed to undergo prothesis and, instead, lost their first member: **psalmu** → **saume**, **pneuma** → **neume**, **phthisis** → **tesie**, Frankish **hlot** → **lot**. This indicates that Romance prothesis was inherited into French in a limited form, applying only before a syllabically stray **s**.

(15) Romance Prothesis 2 (obligatory)

$$\emptyset \rightarrow V \ / \ \#\_\_\ s'$$

Heterosyllabic initials not subject to (15) simplified in the predicted fashion.

In Italian, prothesis became first restricted to heterosyllabic initials found postconsonantally in close-knit phrases: **in iscuola** but postvocalic **la scuola** and **scuola** in phrase-initial position. As in French, prothesis seems to have affected primarily **s-obstruent** clusters, but occasionally other heterosyllabic initials show its effects: Old Tuscan seems to have had prothesis also before ññ (spelled **gn**) initials, as in **ignudo, ignocco** (Rohlfs 1966: 220). Early Italian, like Proto-Romance, must have maintained, in addition to its limited prothesis rule, the old Latin and Romance rule (4): the s-obstruent clusters surface initially regardless of whether prothesis has applied, whereas other heterosyllabic initials fail to surface without prothesis: Gothic **knohha** → **nocca** 'knot', Old High German **slack** → It. **lacco** 'weak'.

Finally, Prothesis was entirely eliminated from Eastern Romance: as a result, Romanian preserves only those heterosyllabic initials whose syllabic incorporation was provided for by rule (4). One can compare the treatment of clusters resulting from the loss of an initial vowel: **expavimen** → ***kspaimen** → **spaima** 'terror', **ekspono** → ***kspono** → **spun** 'I say' but **fossatum** → **fsatu** → **sat** 'village', **budenie** → ***bdenie** → **denie** 'religious service', **dvornicu** → **vornic** 'mayor' (Nandriş 1966: 173). In Old Romanian, the optional elimination of an initial vowel is automatically followed by the simplification of any resulting heterosyllabic initial (**imparatul** → ***mparatul** → **păratul** 'emperor': **insellare** → ***nsellare** → **şălară** 'they saddled', **ambulamus** → ***amblamu** → **mblamu** → **blam** 'we walk'), unless the initial cluster consists of s-obstruent (**Istambul** → **Stambul**, **ispovedire** → **spovedire**).

Needless to add, neither prothesis nor cluster simplification ever affected the initial obstruent-liquid clusters of French, Italian or Romanian.

5. The causes of stress shift

The preceding sections have shown that the only respect in which the syllable structure of Romance differs from that of Classical Latin is the degree to which complex codas had been eliminated. Complex onsets, on the other hand, remained unchanged from Classical Latin into the individual Romance languages to the present.

I return now to one of the questions raised at the beginning of this paper: why did stress shift rightwards in words of the **integra** class? So far we have only concluded that the most obvious explanation, a shift in syllable boundaries, is not available.

I will suggest in this section that the stress shift was due to a reinterpretation of the Latin stress rule.

Several factors had made it difficult to interpret the relation between syllable weight and stress in Proto-Romance. First the syncope of u between a stop and l created penult stressed sequences in **apícla, aurícla, solíclu** etc. from **apícula, aurícula, solículu**. The syncope rule itself was not fully general so that the stress pattern of **apícla** could not be transparently justified in terms of the interaction between stress and syncope. Items like Romanian **títlu** come from unsyncopated Romance ***títulu**: Romance ***títlu** would have become ***tíclu** and then ***tíkyu** in Romanian. Italian **távola** comes from unsyncopated **tábula**.[18] Cul syncope created a very large class of works of the **auricla** type, in which a surface Cl onset is preceded by penult stress. Practically no Latin words like **lócuples** (with an underlying Cl cluster) had been preserved.[19] Nothing would then prevent speakers from deducing that **all** Cl onsets attract stress on the preceding syllable.

Further, the desyllabification of high and mid vowels in hiatus created a second class of words in which a surface complex onset seemed to attract stress on the penultimate syllable. Words like **camísia** 'shirt', **cicónea** 'stork',

albítia 'whiteness', **exágium** 'weighting, weight', **mancípium** 'property' became **ca.mí.sya** (Rom. **cămáşă**), **ci.có.nya** (Fr. **cigógne**), **al.bí.tya** (Rom. **albeáţă**), **es.sa.gyu** (It. **sággio**). That at least some of the C-Glide sequences resulting from Glide Formation were tautosyllabic in Proto-Romance is again indicated by the fact that the initial consonant of the cluster underwent the voicing reserved to postvocalic onsets: **artemísia** → Fr. **armóise** (with [z]); **camísia** → Fr. **chemíse**; **naúsea** → Fr. **nóise**; **ecclésia** → Fr. **églíse**.[20] Words of the **ca.mí.sya** class represented a second apparent instance of stress occurring on the syllable preceding a complex onset. Moreover, since no medial Cy onsets existed in Classical Latin, no words like ***ca.mi.sya** prevented Romance speakers from generalizing: **all** Cy onsets attract stress on the preceding syllable.

Thirdly, a pervasive but optional rule of Late Latin inserted epenthetic vowels into consonant clusters containing **r**. The Latin evidence for this rule was gathered by Schuchardt (1867):

(16) sacrum → sacerum 'sacre'
Alexandri → Alexandiri
supra → supera 'above'
Petro → Petero
mitra → mitara 'cap'

In some cases, the Romance languages preserved the lexicalized outputs of this epenthesis rule: **mitera** (from **mitra**) occurs in Old Italian (Korting 1907: 659), **setenberes** (from **septembris**) and **enterecen** (from **integrent**) occur in Old Spanish (Menendez Pidal 1950: 197). Epenthesis in **r** clusters was continued as an obligatory rule in Sardinian (Wagner 1907): **umbara** from **umbra** 'shadow', **ainturu** from **adintru** 'inside'.

The joint effect of Cul syncope, Glide Formation and **r**-Epenthesis was to obscure the relation between Latin stress and syllabification. The generalization that stress is assigned to a syllable preceding Cl and Cy clusters could be extended further: almost all Romance consonant clusters would appear to have the same effect, simply because most clusters were heterosyllabic. The tell-tale exceptions that might have prevented such a generalization were precisely words like **ín.te.gra**, in which a Cr onset separates the final from the penult. But, due to the optional epenthesis in r clusters, the analysis of **íntegra** was no longer clear: **gr** surfaced both intact and as **gVr**. The result was that very limited data stood in the way of the following reinterpretation of the Latin stress rule:

(17) Stress the penult if it is followed by a consonant cluster.

What must have been the decisive advantage of an approach based on (17) was the fact that it eliminated the opacity of the relation between Cul syncope, Glide Formation and stress: syncope and Glide Formation can now stand in a transparent feeding relation to the stress rule. I suggest then that the stress shift in **integra** was due to the reinterpretation of the Latin Stress rule as (17).

If this suggestion is on the right track, then one might expect a limited range of idiolectal variation to have existed: individual speakers may have succeeded in reconstructing from the available data the Classical Latin stress rule, and its opaque relation to Cul syncope, Glide Formation and r Epenthesis. From such idiolects one may expect forms like the ones in (18), were the stress did not shift rightwards:

(18)	méretrix	Old. Lomb. méltris
	púllitru	Ital. pulédro but Fr. poutre, Cors. póltro, Sp. pótro
	féretru	Fr. fièrtre
	cólobra	Old Astur. cuélebra (Menendez-Pidal 1950: 165)
	cúlcitra	Ital. cúlcetra, O.Sp. acólcetra

Had the rightwards shift in stress represented the automatic result of a change in syllable structure, there would be no way to explain these vestigial forms: throughout Proto-Romance, genuinely closed penults as in **potéstas** had continued to receive obligatory stress.

6. Speculations on the reversal of the stress shift

Both Italian and Spanish have by now reverted to a version of the Latin stress rule: stress cannot recede beyond a closed penult, but it can beyond an open one, even when such a syllable is followed by a complex onset. Harris (1983: 109) cites Spanish **ó.mi.cron, i.dó.la.tra, fú.ne.bre, múl.ti.ple.** Similar Italian examples are **té.ne.bre, cá.li.bro, múl.ti.plo.** In both languages, as everywhere else in Romance, the loss of vowel quantity distinctions added a further complicating factor to the stress rule: since words like Spanish **esquimál** do not differ from words like **caníbal** one must assume lexically governed rules of extrametricality (as in Harris (1983)) in order to account for both types with the same stress rule. In the case of Spanish, extrametricality involves the final segment: hence **es.qui.mál** with a final heavy syllable, but **ca.ní.ba(l)** with a final light syllable. The stress rule itself is identical to that of Latin: stress a final (non-extrametrical) syllable if heavy, otherwise stress the penult. Similar provisions can be made for Italian, where extrametricality, however, involves syllables rather than segments.

Romanian on the other hand, appears to have preserved intact rule (17): of the approximately 45 words longer than two syllables and containing an obstruent-liquid cluster at the boundary between final and penult, none is attested with antepenultimate accent. Loanwords such as **masácru, algébra, echilíbru, escádra, idiolátru, neútru, sperác lu, porécla** are also stressed in accordance with (17).[21]

Two complementary questions need to be asked: why did Spanish and Italian abandon (17) in favor of a revived Latin stress rule? Why did Romanian preserve (17)? My answers to both are speculative.

I begin with the first question. The suggestion in this case will be that Spanish and Italian lost most of the strings that provided critical evidence for (17). The intervocalic Romance complex onsets were either simplified to a single consonant or underwent gemination and became heterosyllabic. The first type of development affected the Northern Italian, Provençal, French and Spanish sequences **VdrV, VgrV,** which become **VirV.** Of the Romance **Cr** onsets, Spanish preserved only the **dr, br, gr** clusters corresponding to Romance **tr, pr, kr**. Tuscan and Southern Italian preserved some instances of both voiceless-r and voiced-r onsets but many fewer than the full set inherited, due to a wide-spread process of liquid metathesis that turned CVCrV sequences into CrVCV. This process operates throughout the Italian dialects, as Milanese **intrég** from **intégro**, Lombard **cadréga** and Venetian **cadréa** from **catédra**, Lucanian **attrúfu** from ottóbre, Sicilian **triátu** from **teátru** (cf. Rohlfs 1966: passim) indicate.

Gemination eliminated the onset **Cl** clusters of Provençal, French and Spanish: they became **C** first and then geminate λλ. In Italian, the Cl onsets became Cy (perhaps through a stage Cλ); then a distinct gemination rule turned all Cy onsets - including those resulting from older Cl - into CCy heterosyllabic sequences (cf. fn. 20). The same gemination rule affected regularly all **labial-r** onsets and, occasionally, other **Cr** onsets.

The global effect of these sound changes was to eliminate from the surface strings of Italian, French and Spanish most postvocalic complex onsets that continued to motivate (17) as the stress rule. In the absence of such strings, the evidence supporting the generalization expressed by (17) - that stress cannot recede beyond a penult followed by a consonant cluster - could be reinterpreted: one could draw the phonologically more natural conclusion that stress cannot recede beyond a closed penult. It is this conclusion that was incorporated into the modern stress rules of Spanish and Italian.

Romanian, on the other hand, is the most conservative of major Romance languages in its preservation of Latin postvocalic onsets. All non-labial voiced stops and all voiceless stops were preserved intact before **r** (Nandriş 1963: 143). The change of l to y in Cl onsets did not affect the structure of the cluster: the resulting Cy sequence did not geminate to CCy as in Italian or become heterosyllabic, as in Spanish. Simply, one type of complex onset, Cl, was replaced by another, Cy. It is conceivable then that the presence of strings like **u.ré.kye** (from **au.rí.cla**), **pă.dú.kye** 'louse' (from ***pe.dí.clus**) etc. continued to provice evidence for the Romance stress rule. In addition, stress shifted rightwards on the penultimate in many cases where a liquid alone separated the final from the penult

(Puşcariu (1903): **búttula** 'small jar' → **butoára, fébru** 'shiver' → **féu.ru** then **fe.ú.ru** and eventually **fi.ó.ru; cúculus** 'bird sp.' → **cocóru, pópolus** 'people' → **popóru; *sécale** (Väänänen 1967:33) → **secára** 'rye'. This last development is not well understood, but represents perhaps the result of an alternative rationalization of the surface stress pattern inherited from Romance. In any event, the presence of a rule like (19) in the contemporary grammar of Romanian, accounts not only for items like **cocór** but also for **masácru, neútru**, etc.

(19) Stress the penultimate if the final onset
contains a continuant sonorant.

Aside from the cases covered by (19), stress is assigned in contemporary Romanian by a rule essentially identical to that of Latin (cf. Steriade 1984) modified by lexically governed extrametricality provisions, as in Spanish.

7. Earlier syllable shifts

We have seen so far that the onset structure of contemporary Romance languages is identical to that of Classical Latin because it is descended from it. The Romance stress shift in the **integra** class, the only phenomenon that could argue for discontinuity in the history of Latin-Romance onset structure, has been given an alternative explanation.

Two complementary syllable shifts have been eliminated in this way: the shift from **in.te.gra** to **in.teg.ra** and the shift back to **in.te.gra**. Recall now that similar shifts in syllabic division appeared necessary for an explanation of the patterns of syncope and vowel raising in prehistoric Latin: from **con.sa.cro**, a syllabification required by **e-** Syncope, one must shift to **con.sac.ro** in order to explain the prehistoric raising to **consecro**, and then back to attested **con.se.cro**. Without the intermediate stage **con.sac.ro**, it would seem impossible to explain why the short **a** did not raise to high (as in **conficio** from **con.fa.ci.o**).

One detail in the paradigm of Raising Indicates however that this traditional explanation for **consecro** is on the wrong track: Raising did take place before Obstruent-l clusters as if the preceding vowel was in an open syllable. Thus ***veheclom** 'vehicle' became **vehiclum**. If obstruent-liquid clusters had been heterosyllabic at the time Raising took place, one would have predicted ***veheclum**.

As Bernardi Perini (1974) has noted, a simple explanation is available for the vowel height of the **consecro** class, which does not require syllable shifts and which does not run afoul of **vehiclum**. Latin lowers i to e before **r**. This rule affects vowels that have been raised in non-initial open syllables: hence the contrast between **pario** 'produce' but **re-perio** 'find' from intermediate, post-Raising ***re-pirio**. The same rule, if extended to apply across an intervening consonant, will correctly derive **consecro** from intermediate **con.si.cro**. Since the lowering of i to e does not take place before l, the vowel quality in **vehiclum** can now be correctly predicted. An elaborate case for this analysis is made in Bernardi Perini (1974), so I will not repeat here the details of the argument.[22]

8. Conclusion

This investigation has a factual conclusion and a number of methodological ones.

The factual conclusion is that the tautosyllabic status of obstruent-liquid clusters in contemporary Romance languages goes back in a straight line to the earliest reconstructible stages of Italic. The history of Latin-Romance syllable structure emerging from this study is one of remarkable stability: over a period of approximately 25 centuries, onset structure has remained essentially the same; coda clusters, originally allowing [+son][-son] sequences, became restricted over the entire documented history of Latin to one consonant only; the same Stray Adjunction rule - (4) above - has provided for the limitations on heterosyllabic initials from Italic and perhaps Indo-European to contemporary Romanian.

Several points of method have also been established. The independent criteria used here in reconstructing syllabic divisions have validated each other. Thus the principles of exhaustive syllabification and onset maximization have led to the reconstruction of syllabic divisions like **re.tro** on the basis of the observation that clusters like **rtr, ltr, str** were preserved in Romance. The assumption of cyclic syllabification has led to Romance ***ad.runco** contrasting with ***qua.drus**: this contrast has explained a fairly complex pattern of gemination (***arrunco** → Rom. **arunc** but **quadru** → Rom. **codru**) which in turn has provided an independent argument for the tautosyllabic status of obstruent-liquid clusters. The assumption that rules which fail to affect geminates must refer to syllabic boundaries or syllable constituents has helped select one formulation of Romance voicing/spirantization, and this choice has once again led to the conclusion that obstruent-liquid sequences were complex onsets.

One important guideline for reconstruction has not been made explicit so far: sound change proper is the source of only a small subset of the transformations in the surface strings of Latin and Romance. Most of the historical developments reconstructed here are changes in the format of existing rules (e.g. the simplification of Latin Regressive Gemination (7) leading to its Romance version (9)); automatic consequences of the interaction between phonological rules and syllabification (e.g. the change in the syllabification of Latin **sta.blu** → ***sta.wlu** (by (12a) → ***staw.lu** → Rom. **staul**; or the automatic cluster simplification following syncope in Provencal **domesticare** → **domestgar** → **domesgar**); or, finally, instances of restructuring (in the sense of Kiparsky 1965). Belonging to the last category are the Romance stress shift of **íntegra** to **intégra** and the Italian/Spanish reversals of this shift: both shifts are due to the changing relation between a stress rule and the surface strings providing transparent evidence for it.

NOTES

[1] On the assimilation of **rl** to **ll** see section 4.

[2] The quality of the vowel preceding the derived **ll** sequence is not always clear: while **sacellum**, **cerebellum** etc. are the expected results, **pugillus**, **sigillum**, **asellus**, **ampulla** cannot be reconciled with the sound laws or with each other. Some discussion of the facts is provided in Leumann 1977:143.

[3] The statements of Latin writers and grammarians on the location of the accent are collected in Schoell (1876).

[4] "The Latin accent has been extraordinarily tenacious and the accented Latin syllables have remained accented in Proto-Romance essentially without exception." (Meyer Lübke 1890 : 522).

[5] Cf. also Väänänen 1967:34 and Pulgram 1975.

[6] A sample of such onset systems is discussed in Levin 1985 and Steriade forthcoming. See also Devine and Stephens 1977. English, whose onset clusters are the same as those of Latin-Romance, behaves in this respect as a French dialect. Germanic languages like Icelandic, which have been shielded from massive Romance influence, have a radically different onset system (cf. Steriade forthcoming).

[7] Harris (1983) and Reighard (1982) report that **tl**, though not **dl**, may be tautosyllabic in Spanish and Portuguese. The Latin evidence for the assignment of such clusters is discussed in Steriade (1982).

[9] There are two alternative ways in which the cluster **ny** may surface: **co[ni]ciō** and **co[yy]iciō**. The first option results from simplifying the initial **yi** to **i**, as elsewhere in Latin (**a.[yi]s** 'you say' → **a.is**). The second option results from the Regressive Gemination rule discussed in the text. Note that both options can be obtained by simply varying the order between the two rules. If **yi** → **i** simplification applies first, Regressive Gemination is bled, for lack of **y**. If, on the other hand, Regressive Gemination applies first, then a geminate **yy** cluster is created, i.e. a single segment multiply associated to both the onset and the coda of the adjacent syllables. Such structures are generally immune to the application of phonological rules which, like **yi** → **i**, require reference to the syllable (cf. Schein and Steriade 1986). For this reason, Regressive Gemination in **co[yy]iciō** bleeds **yi** → **i**.

[10] Latin geminate clusters are simplified after long vowels, hence **glū.ma** from intermediate **glūm.ma**.

[11] On Latin qu (=kw]) see Devine and Stephens 1977 and Steriade 1982.

[12] A different rule of gemination affects certain liquid clusters of decreasing sonority: **rs**, **ls** become **rr** and **ll** respectively; certain old instances of **ln** become **ll** (cf. Niedermann 1953:141). The progressive direction of this rule, as well as its limited scope identify it as a phenomenon distinct from the one discussed in the text.

[13] I believe that Meyer-Lübke (1890:440) was the only one to have noted, with no further comment, the systematic contrast between regressive gemination in cases like **sub-leonem → sollione** and the absence of it in ***tribla, auricla** etc.

[14] Cf. Rohlfs (1966:39) who notes the fact that pre-liquid obstruents are treated by this rule exactly as if they were prevocalic.

[15] Cf. also Weinrich's (1958:238) comments on the context of the Pan-Romanic phenomena of voicing and spirantization: "wo Kurzkonsonanten intervokalisch variiert werden, da werden sie auch zwischen Vokal und **r**, **l** variiert." Weinrich goes on to point to the connection between the treatment of postvocalic obstruent-liquid sequences and the fact that only obstruent-liquid clusters are permitted after consonant in Romance. See below.

[16] The analysis of the medial clusters **kst kstr** found in **sekstus, ekstra** remains unclear: these are the only instances of inter-obstruent **s** found stem-internally in Latin. In general, a cluster [-son]s[-son] loses its first member: **seks-centī → sescentī** 'six hundred', Opscī → **Oscī** 'Oscans'. See Niederman 1953:159, who claims, on the basis of epigraphic evidence, that the **kst** pronunciation was restricted to the literary language.

[17] This fact was already noted by Weinrich (1958:233): "Es gibt nur Dreiergruppen deren letztes Element ein **r** oder l ist. Das ist das vulgärlateinische 'Drei-Konsonanten-Gesetz'."

[18] It has been suggested (most recently by Otero (1985)) that the cluster resulting from syncope was evenly divided among adjacent syllables: thus **au.ri.cu.la → au.ric.la**. Although other cases have been encountered in which the post-syncope syllabification of a cluster differs from the pre-syncope assignment available to the same sort of cluster

(Prince 1980), considerations internal to Romance invalidate this hypothesis: the stop-liquid clusters resulting from syncope do not differ in their evolution from the inherited stop-liquid onsets. Thus the stop in ***lepre** from **leporem** undergoes the postvocalic voicing reserved for onsets, not Regressive Gemination or the change to **i** to which coda stops are subjected in Western Romance: the result is Fr. **lievre** 'hare', Sp. **liebre.** Compare the identical treatment of an inherited **pr** onset in **coprum** 'copper' → Old Fr. **cuevre,** Sp. **cobre.**

[19] The following are all the Classical Latin words of three syllables or longer in which an underlying Cl onset separates a light penult from the final: **cátaplus, céntuplex, cíniflo, lócuples, múltiplex, quádruplus, óctuplus.** None have been preserved in any Romance language, which makes it doubtful that they were present in Proto-Romance at the time of the stress shift.

[20] Stop-y clusters, however, underwent a later Romance rule of progressive gemination: e.g. **sapiat** → It. **sappia,** French **sache** from ***sappyat** → **sappša** → **sapša** → **sašə,** where the absence of postvocalic voicing indicates that the obstruent portion of the cluster had been geminate when that sound change took place. Similarly **habeat** → ***habyat** → ***habbyat** → It. **abbia,** Rom. **aibă,** where the fact that **b** fails to spirantize by rule (12) is again due to the presence of a geminate **bb** at the time of Romanian Spirantization. On progressive gemination within stop-y clusters see Schein and Steriade (1986). The fact itself that the gemination rule applying to the stop-y clusters is not the Regressive Gemination rule in (9) indicates that they were tautosyllabic, like the obstruent-liquid onsets. Had **VCyV** been divided **VC.yV,** one would have expected ***camisya** to undergo Regressive Gemination to ***camiyya,** as in Latin **dis-yungo** → **diyyungo.**

[21] One observes that French is the major source of loanwords into Romanian: if so, the penult stress in examples like **masácru** (from French **massácre**) may be attributed to the influence of the French source. Note however that words like **ne.ú.tru,** which are loans from Latin **néuter**), not from French (**neutre** i.e. [nœtrə]), cannot be so explained.

[22] When presenting this argument at the LSRL conference, I had been unaware of Bernardi Perini's work.

REFERENCES

Al-Mozainy, H.Q., R. Bley-Vroman and J.J. McCarthy. 1985. "Stress Shift and Metrical Structure." **Linguistic Inquiry**, 16, 135-143.

Bernardi Perini, G. 1974. **Due problemi di fonetica latina**, Edizioni dell'Ateneo, Roma.

Devine, A. and L. Stephens. 1977. **Two Studies in latin Phonology**, Anma Libri, Ann Arbor, Michigan.

Fouché, P. 1966. **Phonétique historique du français**, Paris, Klincksieck.

Halle, M. and J.-R. Vergnaud. (Forthcoming) **An Essay on Stress.**

Harris, J. 1983. **Syllable Structure and Stress in Spanish**, MIT Press, Cambridge.

Harris, J. 1986. The Accentual Patterns of Verb Paradigms in Spanish, to appear in Natural Language and Linguistic Theory.

Hayes, B. 1982. "Extrametricality and English Stress," **Linguistic Inquiry**, 13, 227-276.

Hermann, E. 1917. "Lateinische und Romanische Betonung," **Kuhn's Zeitschrift** 18, 1-2, pp. 102-110.

Hermann, E. 1923. **Silbenbildung im Griechischen und in den anderen indogermanischen Sprachen**, Göttingen, Vandenhoeck und Ruprecht.

Juret, A.C. 1921. **Manuel de Phonétique Latine**, Paris.

Kiparsky, P. 1965. Phonological Change. MIT Ph.D. Dissertation.

Körting, G. 1907. **Lateinisch-Romanisches Wörterbuch**, Ferdinand Schoningh, Paderborn.

Leumann, M. et al. 1977. **Lateinische Laut und Formenlehre**, C.H. Beck, Munchen.

Menendez-Pidal, R. 1950. **Origenes del espagnol**, Espasa-Calpe, Madrid.

Meyer-Lübke, W. 1890. **Grammaire des langues romanes**, H. Welter, Paris.

Meyer-Lübke, W. 1920. **Einfuhrung in das Studium der Romanischen Sprachwissenschaft**, Carl Winter, Heidelberg.

Nandriş, O. 1963. **Phonétique historique du roumain**, Paris, Klincksieck.

Niedermann, M. 1953. **Phonétique historique du latin**, Paris, Klincksieck.

Otero, C. 1985. "From Latin to Romance: the Vowel Systems", ms., UCLA.

Pellegrini, G. 1960. **Appunti di grammatica storica del provenzale**, Libreria goliardica, Pisa.

Prince, A. 1980. "A Metrical Theory of Estonian Quantity," **Linguistic Inquiry**, 11, 511-562.

Pulgram, E. 1975. **Latin-Romance Phonology: Prosodics and Metrics**, Wilhelm Fink Verlag, Munich.

Puşcariu, S. 1903. "Rumänische Etymologien," **Zeitschrift fur Romanische Philologie**, 27, 738-748.

Reighard, J. 1982. "Conditions métriques à deux niveaux," paper presented at the Canadian Linguistics Society, Montréal, 1982.

Rohlfs, G. 1966. **Grammatica storica della lingua italiana e dei suoi dialetti**, Einaudi, Torino.

Safarewicz, J. 1969. Historische lateinische Grammatik. Halle, Niemeyer.

Schein, B. and D. Steriade. 1986. "On Geminates," **Linguistic Inquiry**, 13, 691-744.

Schuchardt, H. 1867. **Der Vokalismus des Vulgärlateins**, Teubner, Leipzig.

Selkirk, E. 1978. "The French Foot," **Studies in French Linguistics** 1,2, 141-150.

Schoell, F. 1876. **De accentu linguae latinae veterum grammaticorum testimonia**, Acta Societatis Lipsiae 6, 1-231.

Steriade, D. 1982. **Greek Prosodies and the Nature of Syllabification**, MIT Ph.D. dissertation.

Steriade, D. 1984. "Glides and Vowels in Romanian", in C. Brugman and M. Macaulay (eds.) **Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society**, 47-65.

Steriade, D. (Forthcoming) **Relative Sonority and Syllabification.**

Väänanen, V. 1967. **Introduction au latin vulgaire**, Paris, Klincksieck.

Vendryes, J. 1902. **Recherches sur l'histoire et les effects de l'intensité initiale en latin**, Paris, Klincksieck.

Vogel, I. 1977. **The Syllable in Phonological Theory; with special reference to Italian**, Stanford University Ph.D. dissertation.

Wagner, M.L. 1907. **Lautlehre der Sardinischen Mundarten**, Halle.

Weinricht, H. 1958. **Phonologische Studien zur Romanischen Sprachgeschichte**, Aschendorf, Munster.

Margarita Suñer

Quantifiers and Direct Object-Doubling

The quantifier-like element **todo** enters into what appear to be two different types of structures when in direct object position, i.e., (1) and (2).

(1) Lo vi todo
'I saw it all'

(2) Vi todo
'I saw everything'

At first, it seems tempting to hypothesize that (1) contains an instance of the modifier **todo** while (2) has **todo** in head position. This partitioning would reflect the fact that a referential clitic (CL) enters into a chain with another element; (1) must have the structure in (3). On the other hand, since there is no CL in (2), **todo** must be in head position.[1]

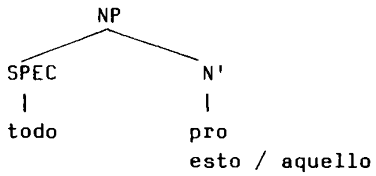
(3) Lo vi [$_{NP}$ todo ec]

However, if the above were correct, one would be at a loss to explain (4) where **todo** cooccurs with a demonstrative.

(4) Vi todo esto / aquello
'I saw all this / that'

What examples of the type in (4) suggest is that even in a sentence like (2) **todo** occupies the modifier position. In essence, I appeal to the null hypothesis: **todo** is always a modifier which might accompany either an overt, or a non-overt head. If correct, this outcome would confirm the hypothesis that empty categories (ec's) parallel the behavior of lexical categories insofar as the language learner has to infer the behavior of the former from the latter (Chomsky 1980 and later work). To state it in slightly different terms, the null hypothesis predicts that wherever a lexical category occurs, a non-lexical one is possible, barring interference from other principles and modules of the grammar. In a nutshell, the claim is that the DO NPs in (1), (2) and (4) have the structure in (5) (first approximation).[2]

(5)



This assessment brings **todo** in line with other specifiers in the language, i.e., like any modifier, it serves as a "nominalizer" (6).

(6) a. la muchacha de la camisa a cuadros
la ec de la camisa a cuadros
the (female) with the checkered shirt

b. compró tres libros
compró tres ec
3sg bought three (books)

The fact that one can say (7a) and (7b) but not (7c) follows independently from the observation that articles are phonological clitics in Spanish, thus they need to cliticize to an appropriate host (the following word or phrase). No such host is available in (7c).

(7) a. Vi todo ec
I-saw all ec

b. Vi a la ec de camisa a cuadros
I-saw the (one) with the checkered shirt

c. *Vi a la ec
I-saw the ec

The main aim of this paper is to substantiate the unitary nature of **todo** by providing evidence from direct object doubling (DO-D). The leading hypothesis is that DO quantified NPs may be doubled when found in an environment in which doubling can take place. This means that the sentence will be grammatical whenever there is feature agreement between the quantified NP and the CL. The advantages of this thesis are multiple: besides rendering a unified analysis of **todo** (vs Jaeggli 1982), it allows one to maintain that doubled DOs are in argument position (vs Aoun 1981 and Hurtado 1984a). Furthermore, it obviates the need to resort to Case mismatch as the explanation for the ill-formedness of some extractions from DO positions (vs Borer 1984).

The data from this study were gathered in Buenos Aires, Argentina, and are representative of the dialect known as Porteño Spanish.

The paper is structured as follows: Section 1. sketches the environment for DO CL-D in Porteño. Section 2. explores doubling possibilities with the quantifier **todo**; and Section 3. refines the structure specified in (5). A conclusion closes the discussion (Section 4.).

1. Direct Object Doubling

Doubling is allowed only when the DO is [+animate +specific]. This means that it is mandatory for all dialects in the context of personal pronouns (8). Moreover, in Porteño DO-D may take place with lexical NPs provided the DO conforms to the required features, cf (9) vs (10). It is important to point out that CL-D is **optional** in this last context, i.e., the sentences in (9) are perfectly grammatical without the CL.

(8) a. **Las** oían a ellas
3pl them heard them
'They heard them'

b. **Nos** quería sólo a nosotros
3sg us loved only us
'S/he loved only us'

(9) a. **La** oían a Paca / a la niña / a la gata
[+anim +spec (+def)]
3pl her listened to Paca / the girl / the cat

b. Diariamente, **la** escuchaba a una mujer que cantaba tangos [+anim +spec (-def)]
Daily, 3sg her listened to a woman who sang tangos

(10) a. *__La__ oían a una niña [+anim -spec (-def)]
3pl her listened to a girl

b. *__Lo__ abracé a alguien [+anim -spec (-def)]
1sg him embraced somebody

c. *__La__ compramos (a) esa novela [-anim +spec(+def)]
1pl it(f) bought that novel

Another point of interest is that when the CL appears non-doubled, i.e., when it enters in a chain with an **ec**[3], although the **ec** must still be interpreted as [+specific] (because the CL must refer to something previously mentioned in the discourse, i.e., the CL acts as a definitizer in this context), the **ec** could be [±animate] (11).

(11) a. Sí, **la** oigo **ec** (where **ec** = e.g. **la niña** 'the girl'or **la música** 'the music')

b. Ya **los** trajeron **ec** (where **ec** = e.g. **los niños** 'the (male) children' or **los paquetes** 'the packages')

The above provides us with the necessary background to be able to return to the behavior of the quantifier **todo** and its variants.

2. Quantifier Doubling

Given the hypothesis that doubling is permitted only in those instances in which there is agreement in features between the DO and the CL, it becomes obvious why a CL is possible in (1) but not in (2). As the translations try to capture, and despite surface appearances, one is dealing with two different variants of **todo**: in (1) **todo** bears the features [+masc-pl] in addition to [+specific] (because of its head); thus, it is compatible with CL cooccurrence. On the other hand, in (2) **todo** is [-count -gder] (cf. Ojeda

1984), an instance of what traditional grammarians classify as neuter gender (Bello 1970, among many others). It is an accident of Spanish that these two variants have the same lexical shape. However, in (4) the relevant features become transparent since **esto/aquello** 'this/that' can only be interpreted as [-count -gder] due to the fact that their corresponding masculine variants are **éste/aquél**. Therefore, the reasons for the ungrammaticality of (12) with CL-D are readily explained: since **eso/aquello** are [-count -gder], they are necessarily **[-animate]**, thus, (12) does not conform to the environment for CL-D because the double is not [+anim +spec]. Contrariwise, since the doubles in (13) carry the necessary features, the examples are grammatical.

(12) a. ***Lo** vi todo eso
lsg him saw all that

b. ***Lo** voy a llevar todo esto / aquello
lsg him am-going to take all this / that

(13) a. (Aún no puedo creer que) **nos** condecoraron a todas nosotras.
(I still can't believe it that) 3pl us decorated all us

b. Ya **los** habían presentado a todos ellos cuando...
3pl already them(m) had introduced them all when ...

Crucially, since the sentences in (13) contain a personal pronoun, the CL is mandatory in **all** dialects of Spanish, cf. the ungrammaticality of (14).

(14) a. *Condecoraron a todas nosotras
3pl decorated all us

b. *Ya habían presentado a todos ellos
3pl already introduced all them

In other words, the examples in (14) are universally ungrammatical in Spanish for the same reason that those in (15) are: personal pronouns must obligatorily cooccur with a CL in all dialects.

(15) a. *Condecoraron a nosotras
3pl decorated us

b. *Ya habían presentado a ellos
3pl already had introduced them(m)

The implication is that **it is the features of the DO phrase which determine the presence / absence of the CL.** The presence of **todo** and its variants plays a relatively minor role: as any other specifier in Spanish, it must agree with its head. Since the "neuter" **todo** is used only in non-specific contexts, CL-doubling is always barred.

And just as one can find grammatical sentences in which **todo** cooccurs with pronominal elements in the head position of the DO NP (13), one can find grammatical instances in which it cooccurs with lexical nouns (16), provided, of course, that the lexical element conforms to the requirement for CL-D, i.e., that it is [+anim +spec].

(16) a. Durante 3 horas tuve la paciencia de escuchar **las** a todas esas ninas
During 3 hours 1sg had the patience of listening to them(f) all those girls

b. Ya **los** había liberado a todos sus esclavos cuando ...
3sg already them(m) had freed all his slaves when ...

On the other had, the sentences in (17) are ungrammatical because the quantifier-like elements (in bold) occur in a DO-NP which does not adhere to the conditions for CL-D, i.e., the DOs are interpreted as [-spec].

(17) a. *No **la** escuché a **ninguna** chica
1sg did not her listen to no girl

b. *Ya **las** vendió a **algunas** esclavas
3sg already sold them(f) some slaves

Moreover, the hypothesis being developed is also capable of accounting for the grammaticality judgments in instances in which the DO is a relative clause (RC). The grammaticality of the sentences in (18) follows automatically from the fact that the [+lexical] head of the RC is [+anim +spec], while the ungrammaticality of those in (19) is a consequence of the head not having the required features.

(18) a. Aunque **las** invité a [$_{NP}$ todas [las ec] [$_{S'}$que estaban en clase]], sólo aparecieron tres
Even though lsg them(f) invited all the (ones) who where in class, only three appeared

b. Ya **las** llamó a [$_{NP}$ todas [las niñas][$_{S'}$que cantan en el coro]]
3sg already them(f) called all the girls who sing in the choir

(19) a. ***Los** escuché (a) todos los discos que me prestaron [-anim]
lsg them(m) listened to all the records that 3pl lent me

b. ***Los** escuché (a) todos los ec que me prestaron [-anim]
lsg them(m) listened to all the (ones) that 3pl lent me

However, given the reasoning above and recalling that (1) - repeated as (20) for ease of exposition - is grammatical because the head position is empty, what accounts for the ungrammaticality of (21)?

(20) **Lo** vi todo ec (**todo** = [+m +sg])
'I saw it all'

(21) ***Lo** escuché todo lo que conseguí (**todo** = [-count -gder])
lsg it listened to all that lsg got

If as explained before, (20) is well-formed because in a non-CL doubled structure the **ec** must be [+spec] but can be either [±anim], shouldn't (21) be grammatical? The answer is a firm no. The reason is that while in (20) the lower NP position (see Section 3) is empty, in (21) this position is partially filled. That is to say, it will immediately be shown that (20) is an instance of a CL ... **pro** chain but (21) is an example of CL-D.

Consider that apriori, two plausible analysis for (21) emerge: (22a) where the **lo** occupies the determiner position before the non-lexical head of the relative, in a way parallel to (18a) (cf. Plann 1980), or (22b) where **lo** forms part of the relative pronoun (cf. Luján 1980).

(22) a. lo escuché $[_{NP}$ todo $[_{NP}$ lo ec] $[_{S}$ [que$_{i}$]
conseguí t_{i} ...

b. lo escuché $[_{NP}$ todo $[_{NP}$ec] $[_{S}$ [lo que$_{i}$]
conseguí t_{i} ...

Although (22a) will prove to provide the correct structure, either analysis renders an ill-formed sentence. The first, i.e. (22a), because, due to the obligatory agreement between heads and determiners, the non-lexical head must agree with the [-count -gder] and therefore, [-anim] **lo**. Hence, the features of the DO are such that CL-D is ruled out. The second, i.e. (22b), because, since relative pronouns agree in features with their antecedents in Spanish, the empty NP cannot be [+anim +spec], thus, the CL is barred again. Consequently, (21) can never be grammatical.

Is there any way to choose between the two analyses for **lo que**? The fact that **todo** by itself cannot serve as the antecedent of a relative clause (Ramsey 1956, 166-167), but needs to be followed by a determiner (23), suggests that in a parallel fashion, **lo** should be considered the determiner of the empty head antecedent (24), and not part of the relative pronoun in COMP. In brief, (22a) displays the correct constituency.[4]

(23) a. todos *(los) que han venido ...
all(m) (the (ones)) who have come ...

b. todas *(las) que compraste ...
all(f) (the (ones)) that 2sg bought ...

(24) Tengo todo *(lo) que quiero
1sg have all (the) that 1sg want

Finally, notice that (25a) without the CL is a grammatical sentence because nothing interferes with the head being characterized by the features [-cont -gder]. It has the structure in (25b).

(25) a. Escuché todo lo que conseguí
lsg listened to all that that lsg got

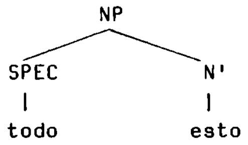
b. escuché $[_{NP}$ todo [lo ec] $[_{S'}$ que conseguí ..

In sum, a unified treatment which analyzes **todo** always as a modifier/specifier explains the distribution of the data in a consistent and natural way, while it confirms the hypothesis that it is the features of the doubled element that determine the grammaticality of CL-D structures.

3. Constituent Structure of QPs

The issue here is what the structure for quantified DOs in a chain with a CL should be. Previously, the structure in (5) - repeated as (26) - was advanced.

(26)

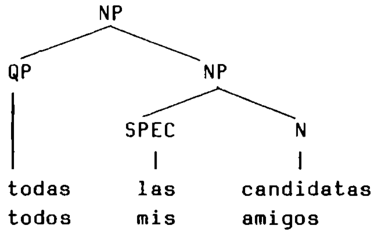


Nevertheless, the existence of examples like those in (27) where the head noun cooccurs with other determiners indicates that the structure in (26) is insufficient. Rather the one in (28) appears to be closer to the correct one. However, note that (26) is still necessary for examples of the type in (29) which bar other specifiers.

(27) a. Aunque ya terminó de entrevistar**las a todas las candidatas,** aún no se decidió por una
Even though 3sg already finished interviewing them(f) all the candidates(f), 3sg has not yet made a decision

b. **Los** encontré muy viejos a **todos mis amigos**
lsg them(m) found very old all my friends

(28)



(29) a. **Todo hombre** debe proceder así
Every man must act that way

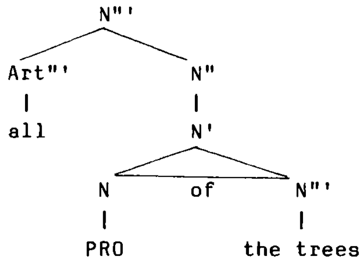
b. Sé que **toda obra importante** va a tener sus opositores
I know that every important work (of art) is going to have detractors

A possible problem for the tree in (28) is that it is in violation of the version of X-bar theory which upholds (30) as valid (for non-adjunctions).

(30) $X^n \rightarrow \ldots X^{n-1}$

However, although (30) may be a desirable way to constrain X-bar theory, it might prove to be too restrictive, given the fact that natural languages are recursive (see Radford 1981:103ff for relevant discussion and references). In addition, there is some evidence that (28) should be a permissible structure in Spanish. Within his three-level approach to X-bar theory, Jackendoff (1977, Chapter 5) divides specifiers into two main groups: those which are N^3 specifiers and those which are N^2 ones. **All,** definite determiners and demonstratives, as well as **which** belong to the first group (i.e. N^3). Transposing Jackendoff's classification to Spanish, note that in (28) two N^3 specifiers cooccur; this is an indication that the structure contains two N^3 since cooccurrence restrictions bar two N^3 specifiers from determining the same N. Jackendoff proposes the structure in (31) for the English phrase **all of the trees.**

(31)



The equivalent phrase in Spanish **todos de los árboles** is ungrammatical because of the presence of **de** 'of'. Since **de** is the grammatical formative used to signal partitiveness, I interpret the ill-formedness of the sequence as an indication that **todo** and its variants do not participate in partitive constructions. If this is correct, (28) would be the representation for Spanish while (31) would be the one for English. However, the final choice between (28) and (31) is not as important as the understanding that the quantifier **todo** enters into the two types of structures exemplified by the sentences in (29) - where the relevant phrase receives a [-specific], i.e. general interpretation - and (27) where the pertinent phrases are read as [+specific] (and definite). This is the crucial distinction for CL-D, because it is only in the latter case that the environment for CL-D materializes.[5]

In view of (28), i.e., taking into account that this structure contains two N^{max}, is there any motivation for considering that the underlined phrases in (27) form a single constituent? The placement of the personal **a** suggests that there is. Note that independently of CL-D, if the DO is [+anim], this **a** materializes before the entire underlined phrase. Any attempt to place the **a** after the quantifier leads to ungrammaticality (32).

(32) a. ***los/Ø** conozco todos **a** los profesores
1sg them/Ø know all **a** the professors

b. ***las/Ø** despidieron todas **a** las tres secretarias
3pl them(f)/Ø fired all **a** the three secretaries

Hence, one must conclude that the quantifier plus the lower NP in (28) form a single constituent. Additional support for this conclusion is found when recalling that whenever the lower NP in (28) is non-lexical, the only indication as to the [±animate] feature of the ec is given precisely by the personal **a** placed in front of the quantifier. Thus, in (33) the absence of a shows that the DO is [-anim], while its presence causes the DO to be interpreted as [+anim], as, for example, in the times of slavery.

(33) Las vendieron (**a**) todas **ec**
3pl them(f) sold (**a**) all

In brief, even acknowledging that (28) violates the postulates of a particular version of X-bar theory, the evidence reviewed points to this tree-type as the one which correctly characterizes the structures under consideration.

4. Conclusion

The data reviewed confirms the initial hypothesis that quantified DO-NPs may be doubled in Porteño Spanish, provided they comply with the environment in which DOs may double, i.e., whenever there is feature agreement between the CL and the Q-NP. In addition, this result lends support to a unitary treatment of **todo** as a specifier element. Moreover, on the assumption that quantifier-like elements like **todo** undergo Quantifier-Raising at LF, examples of the type in which the CL doubles a Q-DO phrase - (13), (16), (18), and (27) - demonstrate that it is possible to extract from CL-D DO position. This outcome has two direct consequences: first, since this extraction renders grammatical results, the doubled DOs must be in argument position (vs Aoun's 1981 and Hurtado's 1984a claims). Second, no Case mismatch between the CL and the DO needs to be resorted to to explain ungrammatical examples (vs Borer 1984).[6]

NOTES

[1] Jaeggli 1982 argues for the double nature of **todo**; see specifically pp 70-71 and 78-79.

[2] Additional evidence for this comes from cross-dialectal data. In Quiteño Spanish (Ecuador), the DO CLs **lo(s)** and **la(s)** are not used in colloquial speech; thus, an example like **Vi todo** is interpreted either as 'I saw it all' or as 'I saw everything'. (I am grateful to M. Yépez for this observation).

[3] The presence of the **ec** is mandated by the Projection Principle which essentially states that the three levels of syntactic representation (LF, S- and D-structure) are projections of lexical properties, i.e., projections of the subcategorization and theta-marking requirements of lexical items as determined by the lexicon.

[4] It might be interesting to speculate as to why the quantifier **todo** cannot cooccur with an empty NP to antecede a relative clause (as in (22b)). The reason might be that the lower NP would be interpreted as a definite personal pronoun and it is a well-known fact that these pronouns cannot serve as antecedents to restrictive relative clauses, irrespective of the presence of **todo** (i).

(i) *Todos nosotros / *Nosotros que estamos de acuerdo lo acompañaremos
All of us / We who are in agreement will-accompany him

[5] Although the preferred place for a quantifier like **todo** is prenominal, it can also appear post-nominally (i). This alternative placement is more common with pronouns. Post-nominal position is concomitant with heavier stress.

(i) a. Nos ayudaron a todos nosotros / a nosotros todos
3pl us helped all us / us all

b. Los admiro a todos ellos / a ellos todos
1sg them(m) admire all them / them all

c. ?Las entrevistaron a las actrices todas
3pl them(f) interviewed the acresses all

[6] This paper is part of a much more comprehensive study on CLs in Porteño entitled "The Role of AGR(eement) in Clitic-Doubled Constructions."

REFERENCES

Aoun, J. 1981. The formal nature of anaphoric relations. MIT dissertation.

Barrenechea, A.M. and T. Orecchia. 1977. La duplicación de objetos directos e indirectos en el español hablado en Buenos Aires. In: J.M. Lope Blanch (ed.), Estudios sobre el español hablado en las principales ciudades de América. México, DF: Universidad Nacional Autónoma de México.

Barwise, J. and R. Cooper. 1981. Generalized quantifiers and natural language. Linguistics and Philosophy 4:159-219.

Bello, A. 1970. Gramática de la lengua castellana. Rev. by J. Cuervo and N. Alcalá-Zamora y Torres. Buenos Aires: Sopena.

Borer, H. 1984. Parametric Syntax. Dordrecht: Foris.

Chomsky, N. 1977. On WH-Movement. In: P. Culicover et al. (eds.) Formal Syntax. NY: Academic Press.

Chomsky, N. 1980. On Binding. L.I. 11.1-46.

Chomsky, N. 1981. Lectures on Government and Binding. Dordrecht: Foris.

Chomsky, N. 1982. Some Concepts and Consequences of the Theory of Goverment and Binding. Cambridge, MA: MIT Press.

Demonte, V. 1977. Partitives and the internal structure of NP specifiers in Spanish. Cahiers Linguistiques D'Ottawa 9. Dept. of Linguistics, Univ. of Ottawa.

Gili y Gaya, S. 1973. Curso superior de sintaxis española. Barcelona: Vox.

Givón, T. 1976. Topic, pronoun and grammatical agreement. C. Li (ed.), Subject and Topic. NY, Academic Press.

Heim, I. 1984. Where does the Definiteness restriction apply? Evidence from the definiteness of variables. Unpublished Univ. of Texas at Austin paper.

Higginbotham, J. and R. May. 1981. Questions, quantifiers and crossing. The Linguistic Review 1: 41-80.

Hurtado, A. 1984a. Clitic chains. Unpublished Simon Fraser Univ. paper.

Hurtado, A. 1984b. On the properties of LF: Cornell Working Papers in Linguistics, 5.

Jackendoff, R. 1977. X Syntax: A Study of Phrase Structure. Linguistic Inquiry Monograph 2. MIT Press.

Jaeggli, O. 1982. Topics in Romance Syntax. Dordrecht: Foris.

Koopman, H. and D. Sportiche. 1982/83. Variables and the Bijection Principle. The Linguistic Review 2. 139-160.

Luján, M. 1980. Sintaxis y semántica del adjectivo. Madrid: Cátedra.

Ojeda, A. 1984. A note on the Spanish neuter. L.I. 15. 171-173.

Plann, S. 1980. Relative clauses in Spanish without overt antecedents and related constructions. Univ. of California Publications in Linguistics, vol. 93.

Radford, A. 1981. Transformational Syntax. Cambridge: Cambridge University Press.

Ramsey, M. 1956. A textbook of Modern Spanish. NY: Holt, Rinehart and Winston.

Real Academia Española (RAE). 1974. Esbozo de una nueva gramática de la lengua española. Madrid: Espasa-Calpe.

Roldán, M. 1971. The double object constructions of Spanish. Language Sciences 15: 8-14.

Selkirk, E. 1977. Some remarks on noun phrase structure. In: P. Culicover et al. Formal Syntax. NY: Academic Press.

Suñer, M. 1982. Big PRO and little **pro**. Unpublished Cornell Univ. paper.

Suñer, M. (1986). The Role of AGR(eement) in Clitic-Doubled Constructions. Unpublished Cornell Univ. paper.

Marie-Thérèse Vinet

Implicit Arguments and Control in Middles and Passives

In recent years, the study of implicit arguments[1] has become a fertile ground for research (cf. Roeper 1983, Rizzi 1985, Williams 1985, Roberts 1985, Jaeggli (to appear), among others). One of the interesting aspects of this area of research is the fact that it converges with the study of control, enlightening what has long been considered one of the murkiest areas of government-binding syntax.

We will examine certain problematic facts for an analysis which assumes that 'only syntactic thematic roles allow control' (Keyser and Roeper 1984). Rather, we will argue, following Jaeggli (to appear), that control with implicit arguments belongs to an independent level of linguistic representation different from 'argument control' which applies at S-structure.

If has often been observed that there are differences in English with control structures in middles and passives. Recent analyses by Baker (1985) and Jaeggli (to appear) have argued quite convincingly that the external theta role is assigned to the passive suffix in all those cases where the 'by-phrase' is optionally or obligatorily missing. With English middles constructions just as with 'able-adjective' constructions and lexical passives, it is well known that a lexical process deletes the external theta role, therefore eliminating a control of the PRO in the embedded clause (cf. Zubizarreta 1985, Jaeggli (to appear)). Nevertheless, we will illustrate that control can become "visible" under certain circumstances within these last structures.

This paper is organized as follows: In Section I we investigate the proposal offered by Roeper (1983) and Keyser and Roeper (1984). We demonstrate in 1.1 that their analysis cannot offer a general and full explanation of the facts. We present cases that are inconsistent with their leading idea which links control to an external mapping phenomenon. These cases are based on data from French and English. We then try to elucidate the facts which suggest that middle and passive constructions sometimes exhibit certain semantic restrictions linked to the adjunct with respect to control. In Section 2 we bring further support to an analysis developed by Williams (1985) which maintains that implicit arguments can serve as antecedents only for arguments, not for adjuncts.

1. A proposal for the analysis of implicit arguments

Much of the analysis in Roeper (1983) relies on the assumption that only syntactic thematic roles allow control. In the same way, Keyser and Roeper (1984) have proposed a form of parametric variation between English and Italian based on the assertion that the notion "implicit agent" is different in middles and in passives. Facts about control follow from this assumption, in their view, since English exhibits a contrast among passives, middles and ergatives:

(1) a. Bureaucrats were bribed to keep them happy
    b. *Bureaucrats bribe easily to keep them happy

(2) a. The buildings were burned to collect insurance
    b. *The buildings burned to collect insurance

As is well known, middle constructions are generated in the grammar by the syntactic rule move- . Following an analysis proposed by Belletti (1983), in the Italian middle sentences the clitic **si** absorbs the Objective Case and bears the thematic role of Agent. It is then assumed that the clitic **si** is generated in the base under the INFL node. Keyser and Roeper (1984) therefore propose that the grammar of English would have an abstract clitic **si** which would also absorb Case and the Agent theme. The basic idea is to give an account of the implied agent which is intuitively found in middles but not in ergatives. In order to explain the contrast observed in (1), they adopt Roeper's distinction between an internal or morphological mapping and an external or syntactic mapping. The main outline of this approach is the following: "the thematic role becomes ineligible for control where there is internal mapping". The absence of control with the English invisible **si** appears to have an explanation within that framework. However, as will be developed in 1.1, some data cannot be explained under this hypothesis.[2]

1.1 On some inconsistencies with the above proposal

Let us consider first some facts from French which exhibit, like Italian and Spanish, a clitic **se** on surface. This clitic also absorbs an Objective Case and bears the thematic function of Agent as argued by Belletti (1983, section 4.2). However, control phenomena do not appear to display exactly the characteristics predicted by Keyser and Roeper's analysis.

It can be observed that control applies quite freely with French middle constructions. Different types of adjuncts can indeed occur with these predicates as illustrated in the following examples:

(3) a. Ce livre se$_i$ lirait facilement PRO$_i$ en sirotant une boisson fraîche
'This book would read easily while sipping a fresh drink'

b. Cette revue se$_i$ feuillette bien PRO$_i$ après avoir mangé
'This magazine reads well after having eaten'

Control can also apply from topic position in a variety of French[3], as in the following example in (4). One of the characteristics of this construction is that there is no c-command relation between the controller in the matrix clause and the controlled PRO in topic position.

(4) PRO$_i$ Avoir l'argent pour le faire, les messages se$_i$ transmettraient vite en pas pour rire (Dialectal French)
'To have the money to do it, messages would transmit very quickly'

However, adverbial purpose clauses cannot appear with some French middle constructions, as in the following examples:

(5) a. *La voiture se conduit vite pour arriver à temps
'The car drives quickly to arrive on time'

b. *Le plancher se lave facilement pour sortir le soir
'The floor washes easily to go out at night'

c. *Le bateau se coule facilement pour ramasser les assurances
'The boat sinks easily to collect insurance'

Apparently, the equivalent examples in Italian are also ruled out. Nevertheless it can be observed that not all purpose clauses are ruled out when embedded into middle constructions. It is perfectly acceptable with one type of middle construction with a dislocation, a construction which is not found in Spanish nor in English:

(6) a. Un bateau, ça se coule facilement pour ramasser des assurances
'A boat, (that) sinks easily to collect insurance'

b. Un chocolat chaud, ça se boit facilement pour se réchauffer l'hiver
'A hot chocolate, (that) drinks easily to warm up in winter'

The difference between (5) and (6) is not very clear. In (5), one may have the impression that the theme of the matrix clause controls the PRO in the purpose clause; this reading is of course ruled out. The presence of the pronominal 'ça' linked to the dislocated NP in (6) seems to somehow facilitate the reading where the agent se controls the PRO of the embedded cluase. We leave this question open pending further research.

Another problem for Keyse and Roeper's analysis is the perfectly grammatical English examples in (7).

(7) a. This blouse washes easily without PRO adding any special product
b. These books sell easily without PRO putting them in the window

Their analysis cannot account for these well-formed sentences. We will consider, just like Zubizarreta (1985), and Jaeggli (to appear) that in middles there is a lexical process in English which deletes the external theta role of the verb. Since the external theta role is missing it cannot participate in controlling the PRO of the adjunct clause. But then, why is the control of PRO still possible in (7)?

In order to answer this question we will look at more complex data where a deleted external theta role gives the impression of sometimes being "visible" for control.

2. Some relevant facts

It is well known that lexical passives and 'able-adjective' constructions are like middle constructions in that they trigger externalisation of the internal argument and, as can be observed in (8), control in the embedded clauses is impossible:

(8) a. *The island is uninhabited to preserve its soil

b. *This paper is not readable to print in this journal

These constructions all have in common the fact that the external theta role can be dispensed with entirely. Since it has been deleted it is therefore normal that we do not find a controlled PRO.

A surprising fact, however, is given in (9), where the presence of **must** linked to **in order to** in the embedded clause produces a well-formed sentence:

(9) a. The island must be uninhabited in order to preserve its soil

b. This paper must be readable in order to print it in this journal

Williams (1985) has already pointed out this situation in his discussion of examples by Chomsky (1982:46):

(10) a. *The books were sold without PRO reading them

b. The books can be sold without reading them

This leads him to postulate that certain adjunct clauses are really complements to the modals (**can ... without**). This situation can also be observed with middle constructions where the pair **must ... in order to** considerably improves the acceptability of the sentence:

(11) a. These cars must sell quickly in order to clear the lot

b. This blouse must wash very easily in order to be able to wear it so fast

Returning to our still unexplained examples in (7) above, we would like to hypothesize, on the basis of examples (9), (10) and (11), that there is an abstract modal linked to **without** which makes the implicit argument "visible". In other words, the without-clause would be a complement of this abstract modal in (7).[4]

Extending Williams' analysis, we would like to suggest that an implicit argument can become "visible" whenever there is a coindexing between an element in the matrix clause and the complement controlled clause. This is exemplified by the pair '**pas assez ... pour**' in French result clauses or 'too Adj... to' in English as in (12b) and (13b):

(12) a. *Cette histoire n'est pas racontable pour aller jusqu'au bout
'This story is not tellable to go right up to the end'

b. Cette histoire n'est pas assez racontable pour aller jusqu'au bout

(13) a. *This bottle is breakable to offer it as a gift
b. This bottle is too easily breakable to offer it as a gift

However, in (12b) and (13b), where is the implicit argument coming from? How should it be interpreted? As we have mentioned earlier, the external theta role has been deleted just like in the case of middles. In order to stay in line with our previous analysis we are forced to look elsewhere for this implicit argument. Moreover, regular adjectives such as 'pretty' or 'expensive' can also appear in this position and they do not imply a reading with a deleted external theta role. We will therefore resort to an analysis presented by Epstein (1984) who claims that the empty complement of certain adjectives is a small **pro** interpreted as a universal quantifier as in:

(14) It is fun (for) x [x to play baseball]

It is clear however, from our examples in (12) and (13) that this implicit argument will become "visible" only when the controlled clause is a complement of some element in the matrix clause.

In (15) we confront a partly different case where there is a reading for PRO in the embedded clause only when the NP in the external argument position has a determiner with a nonquantificational force i.e. a definite or a possessive.

(15) { La } maison est longtemps demeurée inhabitée PRO
{*une} avant de parvenir à la vendre

{'The} house remained uninhabited for a long time
{ *a } before (they) managed to sell it'

The sentence is ruled out with **une** which has a [-specific] and [-definite] reading because **une** is analyzed as an existential quantifier with a narrow scope. This type of quantifier is always excluded in a determiner position of an NP in an external argument position. It is also for this reason that the following sentences are ruled out in French:

(16) a. *un désir de partir
'a desire to leave'

b. *un vif plaisir à vivre à la campagne est unique (Dubé 1985)
'an intense pleasure to live in the country is unique'

As mentioned by Heim (1982), only existential quantifiers with a wide scope are possible in this position, as in: 'Un bon chien obéit à son maître' (a good dog obeys his master).

We would not want to argue that the external theta role is found in the specifier position, as proposed by Roeper (1983) and Chomsky (1984).[5] We believe that Williams (1985) has given convincing arguments against such an analysis. Moreover, it could not give an account of why control is impossible when certain elements are omitted in the matrix clause. Cf. '*la maison est inhabitée avant de parvenir à la vendre' (the house is uninhabited before (they) managed to sell it). In other words there has to be a link between the two clauses and in this case it seems to be with the adverb 'longtemps ... avant de'. This link enables the implicit argument, i.e. the Agent in the specifier position, to become "visible" for control. In Williams' terms, there would then be an association between the Agent in the specifier position and the PRO of the embedded clause.

2.1 Different types of 'control'

In order to elucidate the problem under study we would like to consider other possibilities for control which differ from the regular argument control involving control from an argument position. Williams (1985) has already mentioned S-control or event control where the matrix S itself seems to be the controller of the result clause,[6] as in '[This chalk writes dustless]$_i$ PRO$_i$ to make the teacher's life easier'.

Jaeggli (to appear) identifies a different type of control with implicit argument and calls it 'thematic control'. He shows how thematic control should be a D-structure phenomenon contrary to argument control, which belongs to the S-structure level. He indicates that thematic control does not involve control from an argument position and that it is not possible in passive infinitivals. This is illustrated in (17) [from Jaeggli (to appear)]:

(17) a. The report was carefully prepared to impress the board of directors

b. *The report was carefully prepared to be congratulated by the board of directors

Control into passive infinitivals is also ruled out with our examples (12) and (13):[7]

(18) *This type of bottle is too easily breakable (for) x$_i$ [PRO$_i$ to be invited by the group]

The theta role in the indirect object position in (18) corresponds to some sort of "floating" theta role. It receives a referential content through arb interpretation, a rule which, following Rizzi (1985), can apply either in the lexicon or in the syntax.

To conclude this study on implicit arguments we would like to argue that control with an implicit argument, as discussed in this paper, differs from argument control and should therefore apply at a different level of linguistic representation, namely D-structure.

NOTES

* I would like to thank Osvaldo Jaeggli and Diana Massam for their useful discussion on some of the issues addressed here. Any errors are, of course, my own. This research has been funded in part by a grant from SSHRC Canada (410-85-0105).

[1] The term "implicit argument" is misleading. We argue that an implicit argument does not necessarily correspond to an NP argument position.

[2] Jaeggli (to appear) also rejects Keyser and Roeper's parametric analysis of middle constructions. His refutation is based on data from Spanish.

[3] This construction is found in Quebec French. For a more detailed analysis see Vinet (1985a).

[4] It is unclear why the overt modal **can** cannot appear on the surface with middles. The presence of **can** seems to trigger an epistemic reading and this reading does not imply an agent, as can be observed more clearly with the following ergative verb:

(i) *The buildings can burn without collecting insurance

Since this ambiguity does not show up with **must**, the sentence is not ruled out:

(ii) The buildings must burn in order to collect insurance

[5] Jaeggli (to appear) also argues that nominals such as 'destruction' "contain a PRO in subject position which carries the theta role".

[6] See Jaeggli (to appear) who points out a controversial aspect concerning the S-control analysis.

[7] However, as argued by Jaeggli, control with the empty complement of an adjective would probably be different from thematic control because control into passive infinitivals is usually acceptable in this construction: 'It is fun to be invited first'.

REFERENCES

Baker, M. 1985. Incorporation: A Theory of Grammatical Function Changing. Doctoral dissertation, MIT, Cambridge, Massachusetts.

Belletti, A. 1983. "Morphological passive and PRO-drop". Journal of Linguistic Research 2, 1-34.

Chomsky, N. 1982. Some Concepts and Consequences of the Theory of Government and Binding. Linguistic Inquiry Monograph No. 6, MIT Press, Cambridge, Massachusetts.

Chomsky, N. 1984. Knowledge of Language: Its Nature, Origins and Use. Unpublished ms., MIT, Cambridge, Massachusetts.

Dubé, L. 1985. La variation de l'élément introduisant la proposition infinitive, M.A. thesis, Université de Sherbrooke.

Epstein, S.D. 1984. "Quantifier-pro and the LF Representation of PROarb". Linguistic Inquiry 15, 499-505.

Heim, I. 1982. The Semantics of Definite and Indefinite Noun Phrases, Doctoral dissertation, University of Massachusetts, Amherst.

Jaeggli, O. (to appear). "Passive", Linguistic Inquiry.

Keyser, J. and T. Roeper. 1984. "On the Middle and Ergative Constructions in English", Linguistic Inquiry v. 15, 381-417.

Rizzi, L. 1985. "Null Objects in Italian and the Theory of pro", unpublished paper, MIT, Cambridge, Massachusetts.

Roberts, I. 1985. The Representation of Dethematized Subjects, unpublished doctoral dissertation, USCLA.

Roeper, T. 1983. "Implicit Thematic Roles in the Lexicon and Syntax", ms., U. Massachusetss/Amherst.

Vinet, M.-T. 1985. "Remarque sur deux hypothèses paramétriques", Revue québécoise de linguistique, vol. 15, no. 2, 279-287.

Vinet, M.-T. 1985. "Les relations de contrôle et la grammaire", miméo, Université de Sherbrooke.

Williams, E. 1985. "PRO and Subject of NP", Natural Language and Linguistic Theory 3, 297-315.

Zubizarreta, M.-L. 1985. "The Relation between Morphophonology and Morphosyntax: The Case of Romance Causatives", Linguistic Inquiry, vol. 16, 247-289.